

MANEJO DEL CONOCIMIENTO EN LOS CUADERNOS DE LABORATORIO

OLGA XIMENA GIRALDO PASMIN

7206006

UNIVERSIDAD NACIONAL DE COLOMBIA

FACULTAD DE CIENCIAS AGROPECUARIAS

DEPARTAMENTO DE FITOMEJORAMIENTO

PALMIRA

2011

MANEJO DEL CONOCIMIENTO EN LOS CUADERNOS DE LABORATORIO

OLGA XIMENA GIRALDO PASMIN

7206006

**Trabajo de grado presentado para optar al título de Magíster en CIENCIAS
AGRARIAS**

DIRIGIDO POR:

PhD. ALEXANDER GARCIA CASTRO

PhD. CARLOS GERMAN MUÑOZ PEREA

UNIVERSIDAD NACIONAL DE COLOMBIA

FACULTAD DE CIENCIAS AGROPECUARIAS

DEPARTAMENTO DE FITOMEJORAMIENTO

PALMIRA

2011



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE PALMIRA
FACULTAD DE CIENCIAS AGROPECUARIAS
SECRETARÍA

RESOLUCION No. 162 DE 2011

(Acta No. 021 de diciembre 14)

"Por la cual se otorga la distinción **MERITORIA** al Trabajo de Tesis de la estudiante del Programa de Maestría en Ciencias Agrarias **OLGA XIMENA GIRALDO PASMIN**"

EL CONSEJO DE FACULTAD DE CIENCIAS AGROPECUARIAS, SEDE PALMIRA, en uso de sus atribuciones legales y estatutarias y en especial las otorgadas según el literal d, Artículo 57 del Acuerdo No. 008 de 2008 del Consejo Superior Universitario,

CONSIDERANDO:

Que la Resolución No. 112 de 2010 del Consejo de la Facultad de Ciencias Agropecuarias (Acta No. 010 de Julio 30) estableció los criterios para otorgar mención **MERITORIA** a un trabajo de tesis de posgrado o recomendar ante el Consejo de Facultad la concesión de Mención **LAUREADA**.

Que los jurados Oscar Corcho García, Ph.D., y Leyla Jael García Castro, Ph.D., sustentaron la solicitud de distinción **MERITORIA** ante el Consejo de Facultad de Ciencias Agropecuarias, del Trabajo de Tesis titulado: "**El Manejo del Conocimiento en los Cuadernos de Laboratorio**", realizado por la estudiante de Maestría en Ciencias Agrarias, Línea de Investigación Fitomejoramiento **OLGA XIMENA GIRALDO PASMIN**, bajo la dirección de los doctores Carlos Germán Muñoz y Alexander García.

Que el Consejo de la Facultad de Ciencias Agropecuarias en su sesión No. 021 realizada el día 14 de diciembre de 2011, analizó la solicitud referida y encontró argumentos académicos válidos para otorgar la distinción.

En mérito de lo anterior,

RESUELVE:

ARTÍCULO 1º: Otorgar la distinción **MERITORIA** al Trabajo de Tesis titulado: "**El Manejo del Conocimiento en los Cuadernos de Laboratorio**", realizado por la estudiante de Maestría en Ciencias Agrarias, Línea de Investigación Fitomejoramiento **OLGA XIMENA GIRALDO PASMIN**, bajo la dirección de los doctores Carlos Germán Muñoz y Alexander García.

ARTÍCULO 2º: Entregar la presente Resolución en Nota de Estilo a la estudiante Olga Ximena Giraldo Pasmin.

COMUNÍQUESE Y CUMPLASE

Dada en la ciudad de Palmira, a los catorce (14) días del mes de diciembre de 2011.


MARIO AUGUSTO GARCÍA DÁVILA

Decano


VICTORIA EUGENIA QUINTERO

Secretaria de Facultad

A mi madre.

AGRADECIMIENTOS

A todas las personas que participaron e hicieron posible este proyecto, muchas gracias por su apoyo y enseñanza: Alexander García , Carlos Germán Muñoz (Universidad Nacional, sede Palmira), Proyecto Agrobiodiversidad y Biotecnología del CIAT, Leyla Jael García (European Bioinformatics Institute), Oscar Corcho (Universidad Politécnica de Madrid), Cesar Montaña (Universidad Politécnica de Madrid) y John Bateman (University of Bremen).

Al grupo de profesores de posgrados de la facultad de ciencias agrarias de la Universidad Nacional, sede Palmira por la formación académica.

A mi compañero Hugo Jaimes, por hacer agradables tantas horas de estudio.

A mi familia y amigos, quienes formaron parte de esta etapa y me ayudaron con su compañía, fortaleza, paciencia y amor.

Al comité asesor de posgrados por apoyar esta idea y permitir la realización de esta tesis.

La facultad y los jurados de tesis no se
harán responsables de las ideas emitidas
por el autor.

Artículo 24, resolución 04 de 1974.

TABLA DE CONTENIDO

	Pág.
1. PREFACIO	21
1.1 PREGUNTA DE INVESTIGACIÓN	21
1.2 CONTRIBUCIONES DE ESTA TESIS	23
2. INTRODUCCIÓN	24
2.1 POR QUE SE NECESITA UN LIMS?	27
2.2 QUE BENEFICIOS SE ESPERAN DE UN LIMS?	28
2.3 MANEJO Y REPRESENTACIÓN DEL CONOCIMIENTO EN LOS CUADERNOS DE LABORATORIO.....	35
2.4 BIBLIOGRAFÍA.....	52
3. MANEJO DE CONOCIMIENTO: ANALIZANDO LOS CUADERNOS DE LABORATORIO	60
3.1 INTRODUCCIÓN.....	60
3.2 CARACTERÍSTICAS FÍSICAS DE CUADERNOS DE LABORATORIO ANALIZADOS.....	62
3.3 FORMATO USADO EN LOS CUADERNOS DE LABORATORIO ANALIZADOS..	63
3.4 TIPO DE DATOS ENCONTRADOS EN LOS CUADERNOS.....	70
3.5. METADATOS	86
3.6 COMENTARIOS FINALES.....	101
3.7 BIBLIOGRAFÍA.....	102
4. CONSTRUCCIÓN DE LA ONTOLOGÍA M4L.....	105
4.1 INTRODUCCIÓN.....	105
4.2 PRESENTACIÓN DE LOS PASOS PARA MODELAR LA ONTOLOGÍA M4L	106
4.3 RESULTADOS: DESARROLLO DE LA ONTOLOGÍA M4L PASO A PASO	111
4.4 EVALUACIÓN	143
4.5 COMENTARIOS FINALES.....	145

4.6 BIBLIOGRAFIA.....	147
5. ANOTACIONES LIBRES + METADATOS + ONTOLOGIAS = TAGS4LABS.....	160
5.1 INTRODUCCIÓN.....	160
5.2 MOTIVACIÓN PARA LA CONSTRUCCIÓN DE TAGS4LABS	161
5.3 ESTRUCTURA RETÓRICA DE LOS CUADERNOS DE LABORATORIO	163
5.4 ESTRUCTURANDO LAS ANOTACIONES LIBRES	167
5.5 TAGS4LABS PENSADO COMO UN MODELO DE MULTICAPAS	171
5.6 TAGS4LABS EN ACCIÓN.....	172
5.7 EVALUACIÓN DE TAGS4LABS	176
5.8 COMENTARIOS FINALES.....	178
6. DISCUSION, CONCLUSIONES Y TRABAJO FUTURO.....	182
6.1 DISCUSIÓN.....	182
6.2 CONCLUSIONES.....	184
6.3 TRABAJO FUTURO	186
ANEXOS.....	187
Anexo A. Clases de la ontología general BFO, importadas a M4L	187
Anexo B. Relaciones que hacen parte de OBO_REL, importadas a M4L	189
Anexo C. OBI minimal metadata, ontología de metadatos reusada en M4L.....	190
Anexo D. CONSTRUCCIÓN DE UNA ONTOLOGÍA USANDO PROTÉGÉ.....	191
Como Instalar Protégé.....	191
Como Usar Protégé.	193

LISTA DE FIGURAS

	Pág.
Figura 1. Esquema que representa la planeación del trabajo.	30
Figura 2. Pasos para comprobar la actividad de la enzima Taq DNA Polimerasa.	31
Figura 3. Evaluación de diferentes métodos de extracción de ADN.	32
Figura 4. Anotaciones sobre la evaluación de la respuesta del material vegetal a la técnica de crioconservación.	33
Figura 5. Posibles relaciones entre las entidades: usuarios, tags y recursos. Tomada del Proyecto Morfeo.	38
Figura 6. Captura de pantalla de delicious.	39
Figura 7. Captura de pantalla de connotea.	40
Figura 8. Captura de pantalla de Picasa.	41
Figura 9. Captura de pantalla de facebook.	41
Figura 10. Ejemplos encontrados de como fue anotado el procedimiento de PCR.	42
Figura 11. Mapa conceptual que ilustra una ontología parcial de la técnica de PCR. Tomada de OBI (Ontology for Biomedical Investigation) (Courtot et al., 2008) y Cancer Research and Management ACGT Master Ontology (Anguita et al., 2007).	43

Figura 12. Transición de glosarios a ontologías. Tomado de (Garcia-Castro, 2007).	44
Figura 13. Mapa conceptual que ilustra el fundamento de los vocabularios controlados.	45
Figura 14. Modelo que representa la base de una ontología. Imagen tomada de (Aranguren et al., 2008).	47
Figura 15. Segmento de la ontología Metadata for Laboratory (M4L), desarrollada en este estudio (ver capítulo 4), para ilustrar como protege-2000 (editor de ontologías), presenta los componentes de una ontología.	49
Figura 16. Ontologías del proyecto OBO, que representan el dominio biológico o biomédico desde nivel molecular hasta organismos completos pertenecientes a diferentes taxones. Tomado de (Mungall et al., 2010).	50
Figura 17. Organización jerárquica de las ontologías generales y de dominio, con base al conocimiento que los dos tipos de ontología describe.	51
Figura 18. Características físicas de los cuadernos de laboratorio del proyecto Agro Biodiversidad y Biotecnología del CIAT.	63
Figura 19. Símbolos encontrados para verificar cada paso o cada actividad experimental realizada.	71
Figura 20. Uso de palabras clave.	72
Figura 21. Tipo de listados encontrados en los cuadernos de laboratorio.	72
Figura 22. Forma de presentar los resultados de una actividad experimental.	73
Figura 23. Guías “how-to” encontradas en los cuadernos de laboratorio analizados.	74

Figura 24. Anotaciones relacionadas con “to-dos”, encontradas en los cuadernos de laboratorio.	74
Figura 25. diferentes formas de anotar las fechas en los cuadernos de laboratorio.	75
Figura 26. formula usada en el calculo de diluciones.....	75
Figura 27. anotación hecha acerca de una ruta de archivo electrónico.	76
Figura 28. anotación relacionada con una fuente de información.....	76
Figure 29. tipo de fotografías encontradas en los cuadernos de laboratorio.	78
Figura 30. ejemplos de listados en formato electrónico: a) listado del material vegetal, b) datos fenotípicos importados de libros de campo, y c) listado de cebadores.....	79
Figura 31. resultado parcial presentado en formato digital.	79
Figura 32. Guías How-to, importadas de medios electrónicos a los cuadernos de laboratorio.	80
Figura 33. Representación grafica de vectores. Tipo de figuras encontradas en los cuadernos analizados.....	80
Figura 34. tipo de secuencia parcial encontrada en los cuadernos de laboratorio.	81
Figure 35. A) plantillas de trabajo diseñadas para diferentes actividades experimentales. B) fotografías de geles de agarosa conteniendo anotación manual.....	83
Figura 36. se ilustra el número de veces (4) que fue realizada una PCR bajo las mismas condiciones experimentales.....	84

Figura 37. histograma que representa la frecuencia de registro de cada uno de los 10 subtipos de información manual.	85
Figura 38. histograma que representa la frecuencia de registro de los 6 subtipos de información digital y la frecuencia de registro de los 3 subtipos de información digital con anotaciones manuscritas.	85
Figura 39. Cuadernos de laboratorio no auto-descriptivos.....	87
Figura 40. Cuaderno auto-descriptivo.	88
Figura 41. Presentación de los metadatos que autodescriben un cuaderno de laboratorio y un procedimiento experimental.	100
Figura 42. Definición esquemática de un mapa conceptual.....	108
Figura 43. Pasos para el desarrollo de la ontología M4L.	110
Figura 44. Presentación del propósito de la ontología M4L.	111
Figura 45. Conceptualización y categorización de términos. Segunda etapa del desarrollo de M4L.....	113
Figura 46. Mapa conceptual, que resume las características descritas sobre la ontología OBI.	114
Figura 47. Captura del conocimiento que será representado en la ontología. Tercera etapa del desarrollo de M4L.	117
Figura 48. Mapa conceptual del proceso, cosecha del material vegetal.....	122
Figura 49. Mapa conceptual del proceso, pulverización mecánica del material vegetal.....	124
Figura 50. Mapa conceptual del proceso de extracción de ADN de alta calidad a pequeña escala en plantas.	128

Figura 51. Mapa conceptual que ilustra los componentes de una reacción de PCR.....	131
Figura 52. Mapa conceptual del proceso de la PCR. Imágenes tomadas de University Rochester Medical Center.	133
Figura 53. Mapa conceptual del proceso preparación del gel de agarosa.....	137
Figura 54. Mapa conceptual del proceso carga y corrida de las muestras en un gel de agarosa.....	139
Figura 55. Mapa conceptual del proceso visualización de las muestras en un gel de agarosa y análisis de datos.....	140
Figura 56. Selección de un lenguaje para el diseño de ontologías. Cuarto paso ejecutado para la construcción de M4L.....	141
Figura 57. Selección de una herramienta para el diseño de ontologías. Quinto paso del proceso de diseño de M4L.	142
Figura 58. estructura retórica de los cuadernos de laboratorios.	164
Figura 59. parte de la estructura retórica que soporta la pregunta de competencia 1.	165
Figura 60. parte de la estructura retórica que soporta la pregunta de competencia 2.	165
Figura 61. parte de la estructura retórica que soporta la pregunta de competencia 3.	166
Figura 62. parte de la estructura retórica que soporta la pregunta de competencia 4.	166
Figure 63. estructuración de la anotación semántica y las etiquetas generadas por usuarios.....	168

Figura 64. Soportando la anotación de los registros de laboratorio mediante la AO.	169
Figura 65. Capas de información de Tags4Labs. 1) la provista por el papel en blanco, 2) la provista por anotaciones manuscritas y/o digitales y 3) aquella importada de otros recursos.....	172
Figura 66. iniciar sesión.....	173
Figure 67. Tags4Labs permite conservar un histórico de los cuadernos en papel a manera de imágenes en formato jpg, gif o png.....	173
Figura 68. proceso de anotación en Tags4Labs.	174
Figura 69. proceso de edición de las etiquetas en Tags4Labs.	174
Figura 70. recuperación de información en Tags4Labs.	175
Figura 71. Tags4Labs permite la toma de datos en tiempo real.	176
Figura 72. esquema recuperación de documentos. Adaptado de (Baeza-Yates. and Ribeiro-Neto., 1999).	177
Figura 73. pasos para instalar Protégé. Primera parte.....	191
Figura 74. pasos para instalar Protégé. Segunda parte.....	192
Figura 75. pasos para instalar Protégé. Tercera parte.....	192
Figura 76. pasos para instalar Protégé. Cuarta parte.	193
Figura 77. Pasos para crear una nueva ontología.	194
Figura 78. Selección del lenguaje para el diseño de una nueva ontología.	194
Figura 79. Creación de un URI para una nueva ontología.	195

Figura 80. selección del lenguaje OWL/RDF que mejor se ajuste a la ontología a desarrollar.	196
Figura 81. Escogencia de la interfaz de usuario que se ajuste a sus necesidades.	197
Figure 82. Página principal de Protégé.	197
Figura 83. pasos a seguir para visualizar los términos de una ontología. Primera parte.	198
Figura 84. Pasos a seguir para visualizar los términos de una ontología. Segunda parte.	199
Figura 85. pasos a seguir para visualizar los términos de una ontología. Tercera parte.	200
Figura 86. pasos a seguir para visualizar los términos de una ontología. Cuarta parte.	200
Figura 87. pasos a seguir para visualizar los términos de una ontología. Quinta parte.	201
Figura 88. Pasos a seguir para visualizar los términos de una ontología. Sexta parte.	201
Figura 89. Pasos para importar ontologías disponibles en Internet. Primera parte.	202
Figura 90. Pasos para importar ontologías disponibles en Internet. Segunda parte.	203
Figura 91. Pasos para importar ontologías disponibles en Internet. Tercera parte.	203

Figura 92. Pasos para importar ontologías disponibles en Internet. Cuarta parte.	204
Figura 93. Pasos para importar ontologías desde un archivo local. Primera parte.	204
Figura 94. Pasos para importar ontologías desde un archivo local. Segunda parte.	205
Figura 95. Pasos para importar ontologías desde un archivo local. Tercera parte.	205
Figura 96. Pasos para importar ontologías desde un archivo local. Cuarta parte.	206
Figura 97. crear una clase en Protégé. Primera parte.	207
Figura 98. crear una clase en Protégé. Segunda parte.....	208
Figura 99. crear una clase en Protégé. Tercera parte.....	208
Figure 100. crear una clase en Protégé. Cuarta parte.	209
Figure 101. crear una subclase en Protégé.	210
Figura 102. diferentes niveles en la jerarquía de clases. El término biomacromolécula (Biomacromolecule) es uno de los conceptos más generales de la ontología, (nivel superior). Los términos ADN nuclear (Nuclear_DNA), ADN mitocondrial (Mitochondrial_DNA), ADN complementario (Complementary_DNA), ADN de cloroplasto (Chloroplast_DNA), ARN mensajero (Messenger_RNA), ARN de transferencia (Transfer_RNA) y ARN ribosomal (Ribosomal_RNA) son un ejemplo de las clases más específicas en la jerarquía (nivel inferior).	210
Figura 103. diferentes tipos de propiedades OWL. Tomado de Matthew Horridge, 2011	211

Figura 104. Crear una propiedad de objeto. Primera parte.....	212
Figure 105. Crear una propiedad de objeto. Segunda parte.	213
Figura 106. ejemplo de una propiedad funcional. Tomado de Matthew Horridge, 2011.	214
Figura 107. ejemplo de una propiedad funcional inversa.....	214
Figura 108. ejemplo de una propiedad transitive. Tomado de Matthew Horridge, 2011	215
Figura 109. ejemplo de una propiedad simétrica.	215
Figura 110. pasos para crear una propiedad especificando el dominio y rango.	216

LISTA DE TABLAS

	Pág.
Tabla 1. Tipos de información manuscrita identificada en los cuaderno de laboratorio analizados.	77
Tabla 2. Tipos de información digital identificada en los cuadernos de laboratorio analizados.	82
Tabla 3. Tipos de información digital conteniendo anotación manuscrita identificada en los cuadernos de laboratorio analizados.....	83
Tabla 4. Metadatos asignados a diferentes objetos.	89
Tabla 5. Conjunto de metadatos que soportan la auto-descripción del cuaderno de laboratorio.	93
Tabla 6. Conjunto de metadatos que soportan la auto-descripción de una actividad experimental.....	98
Tabla 7. Presentación del número de conceptos provenientes por ontología. ...	116
Tabla 8. Tipos de anotación provistos por la AO.	170
Tabla 9. Clases de la ontología general BFO, importadas a la ontología M4L...	187
Tabla 10. Relaciones que hacen parte de OBO_REL, importadas a M4L.	189
Tabla 11. Metadatos obligatorios.....	190
Tabla 12. Metadatos opcionales	190

1. PREFACIO

La diversidad de información que los investigadores en ciencias biológicas deben manejar hace compleja las labores de análisis. Los investigadores actualmente entre sus quehaceres diarios almacenan y clasifican información procedente de equipos, bases de datos, además de aquella producida en su entorno inmediato de manera continua por sus colegas. Adicionalmente, los investigadores, a partir de las anotaciones hechas en sus cuadernos de laboratorio, elaboran informes, analizan resultados y comparten datos experimentales. Los cuadernos de laboratorio son mayoritariamente usados aun hoy en plena era digital debido, en parte, a la simplicidad que proporciona el papel en las tareas de diseño, captura, almacenamiento, clasificación y organización de datos. Además, este facilita la conservación del registro histórico y descriptivo tanto de éxitos como de fracasos en los experimentos realizados. La información contenida en los cuadernos de laboratorio no esta estructurada, se maneja de manera manuscrita, no se ciñe a formatos predefinidos y es de obligatorio cumplimiento.

En esta investigación se propone un marco de trabajo, basado en la aplicación de principios de Web Semántico y Folksonomías, para facilitar el manejo de la información consignada en los cuadernos de laboratorio.

1.1 PREGUNTA DE INVESTIGACIÓN

En esta tesis se estudia el problema de manejo de conocimiento en el escenario

de los cuadernos de laboratorio. Específicamente se reporta acerca de la aplicación de principios y prácticas de las folksonomías combinadas con tecnologías del Web semántico en este escenario. El propósito de esta tesis es facilitar la organización, clasificación y el compartir la información consignada en este tipo de documentos. La pregunta general de investigación planteada en esta tesis es:

“Como organizar, clasificar y compartir la información consignada en los cuadernos de laboratorio?”

Para este efecto se analizaron 12 cuadernos de laboratorio facilitados por investigadores del área de Biotecnología del Centro Internacional de Agricultura tropical, CIAT. A partir del análisis realizado se determinaron los tipos de datos consignados, maneras de organizar los documentos, estructura de los documentos, multiplicidad de información presente –aquella que proviene de un recurso externo como un listado, fotografías de un gel, etc.- Se analizó también como el cuaderno de laboratorio constituye un repositorio de conocimiento, datos e información útil para los investigadores. Se estudió de igual manera el uso que los investigadores le dan a estos artefactos para compartir su conocimiento.

Con el propósito de hacer mas manejable la pregunta general de investigación esta se ha fragmentado en:

- Como facilitar el proceso de anotación (tagueo), de la información disponible en los cuadernos de laboratorio siendo estos originalmente escritos a mano alzada?
- Como facilitar el manejo y manipulación de los tags generados para permitir una clasificación dinámica de los documentos así marcados?
- Como complementar los tags así generados con las ontologías disponibles?

El resultado de esta tesis es un marco de trabajo que permite la preservación digital de los cuadernos de laboratorio en un ambiente de software que facilita la clasificación dinámica y la recuperación de información contenida en estos documentos. Se hace posible además el establecimiento de una red social basada en el contenido consignado en los cuadernos de laboratorio.

1.2 CONTRIBUCIONES DE ESTA TESIS

Tres son las mayores contribuciones de este trabajo, a saber:

- Mejoramiento en la comprensión relativa al uso de anotaciones libres o tags, y su aplicabilidad a un escenario de documentos científicos no estructurados.
- Avance en el entendimiento de la relación entre ontologías y folksonomías en ciencias biológicas.
- Ampliación y mayor claridad en el uso de metamodelos existentes en el caso específico de los cuadernos de laboratorio y las folksonomías generadas sobre la información en ellos consignada.

2. INTRODUCCIÓN

El cuaderno de laboratorio constituye un registro completo de todo el trabajo práctico realizado (Kanare, 1985); este contiene toda la información necesaria (operaciones, hechos observados y conclusiones), para que cualquier investigador pueda reproducir el trabajo realizado (Kanare, 1985). Al ser llevado de forma manual, se espera que los cuadernos de laboratorio sean escritos de manera legible y llevados de manera ordenada. Estas condiciones operativas son primordiales; por ejemplo, en casos relacionados con propiedad intelectual el inventor tendría así la capacidad de defender con éxito una patente. Apoyado en un registro claro, preciso y con fechas exactas de las actividades experimentales o procesos de laboratorio que llevó a cabo para la obtención del invento (Kanare, 1985, Rebbeck, 2005).

Un ejemplo que ilustra la importancia de las anotaciones consignadas en el cuaderno de laboratorio lo constituye la situación generada en relación al científico Charles H. Best¹. En diciembre de 1923 se entregó el Premio Nobel en Medicina y Fisiología a los científicos que dos años antes descubrieron la insulina. Sin embargo, no todos los co-descubridores fueron designados por el Comité de los Premios Nobel para recibirlo. Sólo Frederick G. Banting² y John J. R. Macleod³

¹ Charles Best. Wikipedia, the free encyclopedia.

² Frederick Banting. Wikipedia, the free encyclopedia.

³ John James Richard Macleod. Wikipedia, the free encyclopedia.

fueron distinguidos con la nominación; Charles H. Best no fue incluido simplemente porque no pudo reproducir el aislamiento y caracterización del polipéptido. Esto debido a que no contaba con las notas del procedimiento que previamente había ensayado de forma positiva. Situaciones similares ocurren con frecuencia en profesionales dedicados a la ciencia o a la tecnología (Rodríguez-Yunta, 2008).

El uso de los cuadernos de laboratorio también permite acelerar la estandarización de procesos, mantener un control de calidad y obtener resultados reproducibles (Kanare, 1985). Otras características tales como la portabilidad, facilidad de edición, bajo costo, entre otras, han hecho que el cuaderno en papel siga siendo el medio mas utilizado para el registro de datos científicos. Sin embargo, el cuaderno de laboratorio tiene limitaciones tales como las reportadas por Sarini, et al. (Sarini et al., 2004), Lysakowski (Lysakowski, 1997) y Polonsky, et al. (Polonsky et al., 2011):

- Pérdida de información por no usar cuadernos adecuados: la información en papel puede perderse cuando no se usa un cuaderno con un tipo de papel de características físicas mínimas requeridas para ser funcionales en un ambiente de laboratorio.
- Búsqueda manual de la información: a pesar de ser un registro actualizado, las búsquedas automatizadas no son posibles.
- Dificultad para compartir la información: la información almacenada en los cuadernos no puede ser compartida o distribuida fácilmente entre colegas. Fotocopiar es la manera mas practica, pero legalmente no es valido el crear múltiples copias de los registros experimentales consignados en los cuadernos. Además, si la información del cuaderno no esta escrita de manera legible, en ausencia del autor de la información se dificultará la reproducibilidad del trabajo.
- Captura de datos: la captura de datos se hace de manera manuscrita. Por ejemplo, los resultados que soportan el manuscrito muchas veces se

encuentran en archivos electrónicos que deben ser impresos y pegados en el cuaderno. Esta es una practica muy común incluso en laboratorios modernos.

- Capacidad limitada: los cuadernos están limitados por su capacidad en numero de hojas. Esto hace imposible el incluir, por ejemplo, videos. Tampoco se puede manejar el volumen de datos en bruto que producen algunos equipos de laboratorio de alta eficiencia; esta cantidad puede llegar a ser del orden de los terabytes.
- Mayor tiempo en el planeamiento del trabajo: se estima que en los laboratorios que emplean los cuadernos de papel, los científicos invierten mas tiempo en el planeamiento de su trabajo ya que gran parte de ese tiempo se puede atribuir a la escritura y re-escritura de protocolos similares.
- La organización del cuaderno: ésta varia de usuario a usuario y no sigue un estándar. Una alternativa para mejorar este aspecto, es la construcción de una tabla de contenido, sin embargo, este procedimiento es ignorado por la mayoría de investigadores por el tiempo que debe invertirse en ello.

Esta serie de desventajas han sido parcialmente cubiertas por herramientas tales como el LIMS (Laboratory Information Management System). Este es un sistema que permite la adquisición y gestión de la información generada en el laboratorio (Jayashree et al., 2006). Actualmente los LIMS, están al alcance de la mayoría de las economías, ya que los costos de los mismos se ajustan a las necesidades y a los equipos informáticos que posee un laboratorio. Además aun siendo sistemas informáticos sofisticados, suelen ser fáciles de usar y de poner en funcionamiento (Köhl et al., 2008).

2.1 POR QUE SE NECESITA UN LIMS?

Con la llegada de nuevas tecnologías, en los laboratorios se generan gran cantidad de datos, estos deben ser debidamente procesados para la generación de resultados (Pajon et al., 2005). Además, para cumplir con los estándares de calidad, en los laboratorios debe existir un control riguroso en los procedimientos utilizados en la adquisición y procesamiento de los datos; esto con el fin de asegurar tanto la integridad de la información como la validez de las decisiones tomadas (Paszko et al., 2002, Haquin et al., 2008). Teniendo en cuenta lo anterior, un LIMS es de gran ayuda, ya que éste sistema está organizado de manera jerárquica. Es decir, un LIMS puede contener un conjunto de bases de datos capaces de normalizar y procesar la información generada en un laboratorio de biología molecular. A su vez estas bases de datos, pueden contener sub-bases de datos que ayuden al manejo de datos relacionados con muestras biológicas como clones, oligonucleótidos, reactivos, etc. (Jayashree et al., 2006, Hilliard et al., 2001).

De igual manera, las Agencias Reguladoras han resaltado la importancia de validar la información antes de incluirla en un sistema automatizado. Se espera entonces que algunos de los beneficios obtenidos mediante el uso de un LIMS en este campo, control de calidad, incluyan (Paszko et al., 2002):

- Restricción al acceso en la entrada y/o visualización de datos.
- Control automático de la validez de los datos entrados.
- Conexión con los equipos del laboratorio.
- Uso de Códigos de Barras.
- Cálculos automáticos.
- Generación automática de informes y gráficos.

2.2 QUE BENEFICIOS SE ESPERAN DE UN LIMS?

Las ventajas que supone incorporar un sistema de éstas características al laboratorio han sido descritas por Paszko, et al., (Paszko et al., 2002) y Hilliard, et al. (Hilliard et al., 2001):

- Reproducibilidad: se espera que toda la información generada con anterioridad, pueda ser reproducida en cualquier momento y de forma sencilla.
- Seguridad (control de cambios en la información): el sistema debe impedir modificaciones no autorizadas y mantener un historial de los cambios realizados.
- Revisión y visualización de datos: se espera que sea fácil el revisar y visualizar datos.
- Generación de informes: se pueda generar de manera rápida y efectiva.
- Seguimiento de estados: se espera que el LIMS informe a los usuarios sobre el estado de sus proyectos, muestras, resultados, informes, insumos, mantenimiento preventivo de equipos, y cualquier otra tarea que requiera un laboratorio, de manera automática.

Se estima que el incremento de productividad con la implementación de un sistema LIMS, sea entre un 20% y 30%. Los puntos enumerados a continuación influyen directamente en el aumento de la productividad (Hilliard et al., 2001):

- Reducción de entrada manual de datos.
- Eliminación de la revisión manual de datos.

- Planificación del trabajo del laboratorio y generación automática de hojas de trabajo.
- Realización automática de cálculos y gráficos.
- Generación automática y eficiente de informes rutinarios y ocasionales. No dependiente de la complejidad del informe ni de la carga de trabajo del personal de laboratorio.
- Automatización de acciones condicionadas a determinadas situaciones; por ejemplo, emisión automática de informes o mensajes de alerta, cuando se requiera la calibración de instrumentos o equipos del laboratorio.
- Transferencia automática de información desde y hacia otros sistemas de la compañía.
- Integración con software de uso habitual (Procesadores de Texto, Hojas de Cálculo) sin pérdida de seguridad y sin coste en el tiempo.

Si bien un LIMS es un sistema que asiste al investigador en la organización y control del trabajo de laboratorio, así como también en las etapas de los procesos analíticos y de gestión (Paszko et al., 2002), el llevar un cuaderno de laboratorio sigue siendo una práctica común en cualquier tipo de investigación (Lysakowski, 1997). La razón principal para que el cuaderno de laboratorio siga vigente se debe a la facilidad que éste brinda al investigador para generar todo tipo de anotaciones y manejar cualquier tipo de datos (Sarini et al., 2004). En el cuaderno de laboratorio, se captura el historial de una serie de experimentos y el conocimiento de los investigadores; además es el medio a través del cual se comparte de manera explícita el conocimiento allí consignado (Myers et al., 1996).

A continuación se exponen 3 ejemplos que ilustran como los investigadores anotan la información de procesos de rutina como la PCR (del inglés, Polymerase Chain

Reaction), extracción de ADN y de crioconservación. Lo común en estos ejemplos es que en los cuadernos se conserva un historial del proceso de estandarización de cada uno de los procedimientos antes mencionados; mas aún, hasta la fecha ningún paquete de software logra hacerlo de manera practica y eficiente (Polonsky et al., 2011).

- Comprobación de actividad de la enzima Taq DNA Polimerasa (Ejemplo 1). La figura 1 ilustra la planeación del trabajo, en el cual se resumen los ensayos a seguir durante el proceso de verificación de la actividad de la enzima utilizando la técnica de la PCR. Posterior a ello (figura 2), el investigador anota las condiciones usadas en cada experimento, quedando un registro tanto de éxitos como de fracasos para cada ensayo realizado. Esto representa un recurso importante para verificar la reproducibilidad o posibilidad de repetición del trabajo realizado y al mismo tiempo se mantiene un control de calidad.

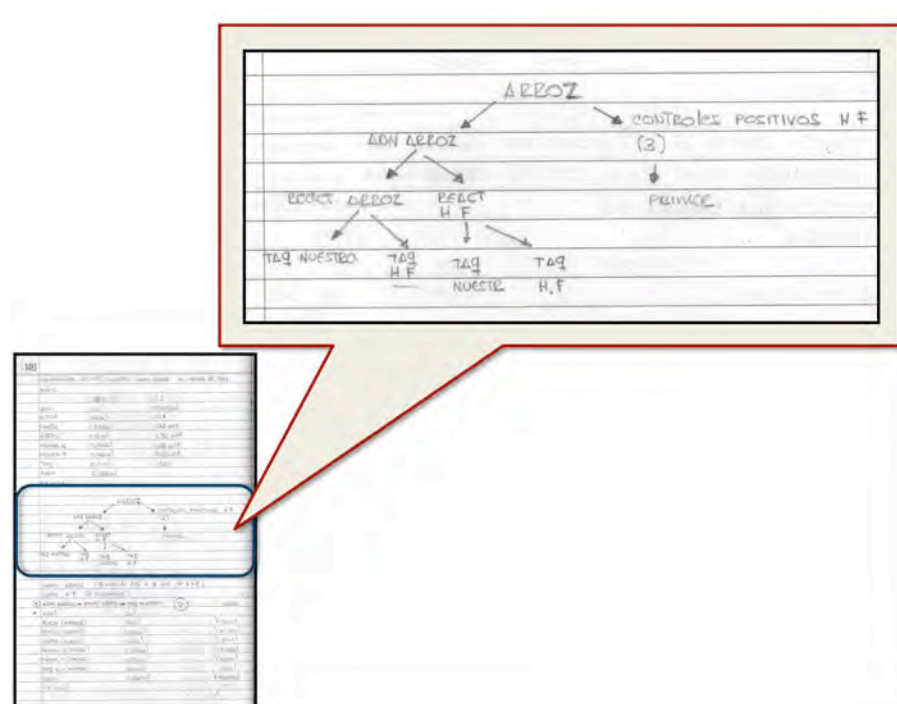


Figura 1. Esquema que representa la planeación del trabajo.

A continuación se ilustra de manera secuencial, los pasos que sigue un investigador para comprobar la actividad de la enzima Tag DNA Polimerasa en un proceso de PCR.

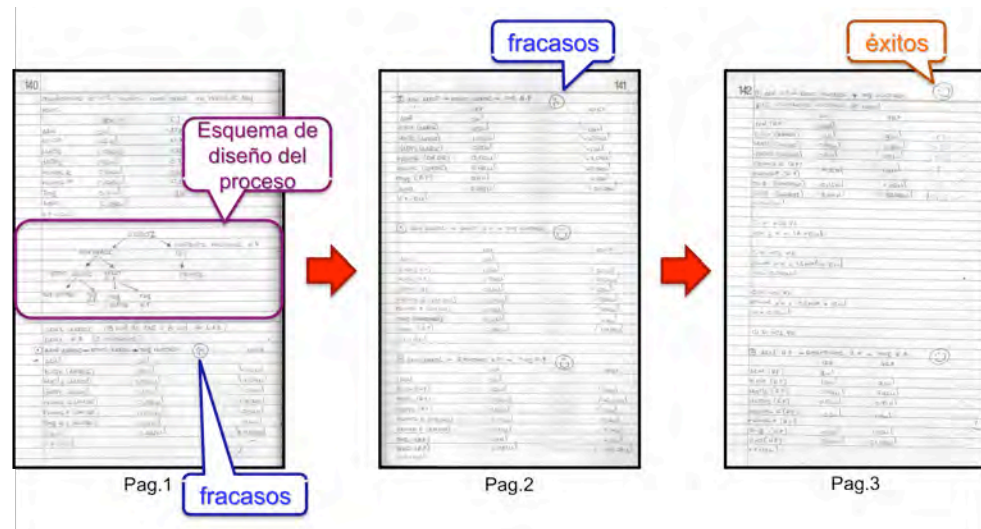


Figura 2. Pasos que sigue un investigador para comprobar la actividad de la enzima Tag DNA Polimerasa.

- Evaluación de diferentes métodos de extracción de ADN (Ejemplo 2). En el campo de la biología molecular, el ADN ha sido el biomaterial mas usado; a partir de éste, se han desarrollado multitud de herramientas tecnológicas que explotan sus propiedades fisicoquímicas para analizar su implicación en problemas concretos. Por ejemplo, desde análisis filogenéticos para detectar similitudes entre taxones, hasta la caracterización de un individuo o grupo de individuos en su respuesta a una característica en particular (Voelkerding et al., 2009). Siendo el ADN el punto de partida para encontrar respuestas en el ámbito molecular, la escogencia de un protocolo de extracción de ADN adecuado es esencial para el desarrollo de un proyecto de investigación.

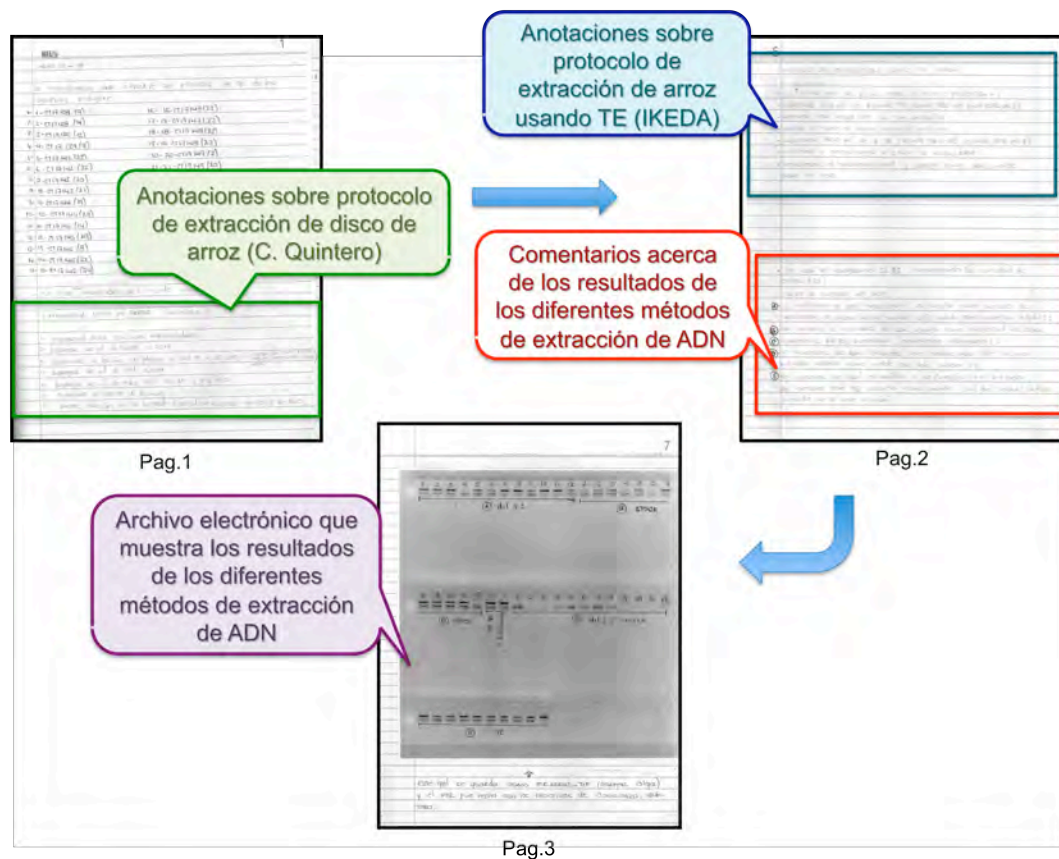


Figura 3. Evaluación de diferentes métodos de extracción de ADN.

En la figura 3 se ilustran las anotaciones realizadas por un investigador para la escogencia del protocolo de extracción de ADN. En este caso el investigador usa como control un numero de muestras de ADN vegetal extraído con un método sugerido por Dellaporta (Dellaporta et al., 1983), y los compara con el ADN vegetal obtenido a partir de otros dos métodos descritos en las paginas uno y dos (figura 3). Finalmente, en la pagina tres (figura 3), el investigador a pegado en su cuaderno una foto del gel de agarosa donde se visualizan las muestras de ADN, lo que le permite hacer la selección del método de extracción de ADN a seguir y al mismo tiempo es la evidencia que soporta su trabajo.

- Experimento de crioconservación (Ejemplo 3). La evaluación de la respuesta del material vegetal a la técnica de crioconservación (herramienta de cultivo in Vitro que permite conservar a largo plazo recursos fitogenéticos), requiere por parte del investigador el mantener un registro detallado de cada uno de sus experimentos. Los mismos investigadores son enfáticos en expresar que el único medio que les permite llevar un registro histórico de sus ensayos es el cuaderno de laboratorio⁴. La figura 4 ilustra las anotaciones hechas por el investigador, en un experimento donde se evalúa la respuesta del material vegetal a la técnica de crioconservación.

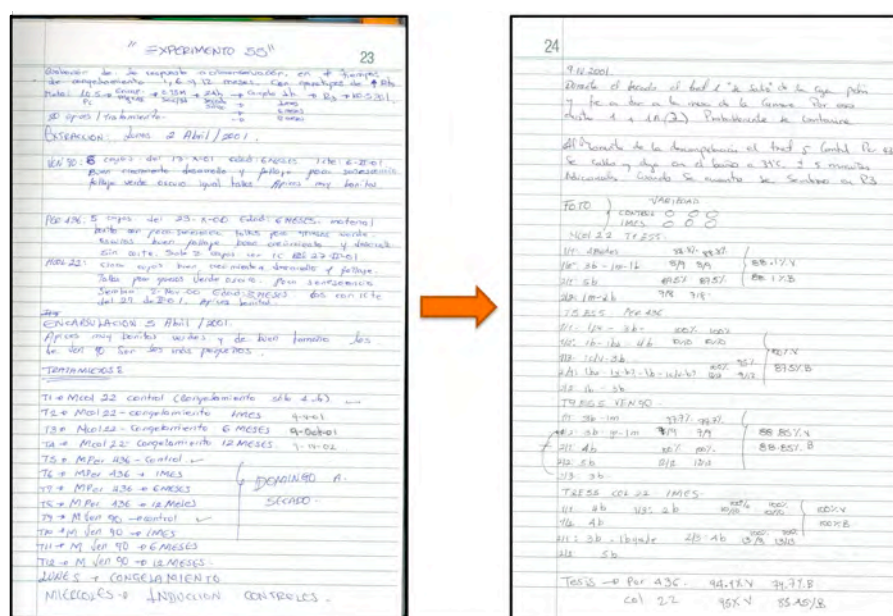


Figura 4. Anotaciones sobre la evaluación de la respuesta del material vegetal a la técnica de crioconservación.

Los ejemplos anteriores, ilustran la importancia de la información que se consigna en el cuaderno de laboratorio y en los archivos electrónicos que cada investigador genera. De igual manera se reflejan algunas limitantes cuando la información no

⁴ ESCOBAR, R. (2008) comunicación personal. CIAT.

sigue ningún estándar como ocurre en los cuadernos de laboratorio; para demostrar esto, se considera el siguiente escenario:

Un investigador que requiera información de un colega ausente de su sitio de trabajo: en este escenario el investigador se enfrenta con dos situaciones. La primera de ellas, es hacer una “búsqueda ciega” en él o los cuadernos del investigador dueño de la información, sin saber a priori si el cuaderno contiene la información que necesita. La segunda es, si logra encontrar lo que busca, ahora debe enfrentarse a la interpretación de la información, que puede estar de manera manuscrita y organizada de acuerdo al estilo personal y experiencia del colega ausente.

De acuerdo a Sarini, et al. (Sarini et al., 2004), y Butler (Butler, 2005), la mayoría de los problemas relacionados con el manejo de la información biológica proviene de la forma en que ésta es inicialmente capturada y almacenada. Los cuadernos de laboratorio electrónicos o ELNs (del inglés, Electronic Laboratory Notebooks), surgen como respuesta a la necesidad de conservar este conocimiento en formato electrónico y de manera estructurada; para facilitar así su acceso, intercambio y reutilización por parte de los miembros del laboratorio. Estos dispositivos de hardware y software facilitan la captura de la información de manera electrónica; además se constituyen en repositorios donde se almacena la información tanto de las notas experimentales como los archivos electrónicos (Polonsky et al., 2011).

Hasta aquí, el almacenar, organizar, descubrir y compartir el conocimiento de forma tal que represente beneficios y productividad en un laboratorio, parece haber sido superado con la llegada de los ELNs, sin embargo, vale la pena preguntarse por qué existen tantos ELNs hoy en el mercado y pocos laboratorios haciendo uso de ellos? (Rebbeck, 2005, Sarini et al., 2004, Lysakowski, 1997, Polonsky et al., 2011, Myers et al., 1996, Butler, 2005, 2008), la respuesta a esta pregunta esta ligada a la manera como se maneja el conocimiento en los ELNs, ya que las personas que actualmente tienen acceso a esta tecnología deben elegir entre: 1) un sistema

flexible, pero la entrada de los datos se hace de manera no estructurada y 2) el sistema es rígido, la entrada de datos se hace de manera estructurada. Estas son las principales razones por las cuales estos sistemas no han sido ampliamente adoptados por los investigadores (Elliott, 2006). Como resultado, una gran cantidad de la información generada durante la investigación se pierde en el tiempo siendo además difícil de recuperar, comprender, analizar y manipular (Butler, 2005).

2.3 MANEJO Y REPRESENTACIÓN DEL CONOCIMIENTO EN LOS CUADERNOS DE LABORATORIO

2.3.1 Conocimiento. Para un mejor entendimiento en relación al manejo y representación de conocimiento, es importante tener claro el termino “conocimiento”. Sin embargo, esta definición no ha sido fácil (Sparrow, 1998, Sandars and Heller, 2006); antecedentes indican que los debates acerca de la naturaleza del conocimiento han sido registrados en la filosofía griega desde el periodo clásico. En esos escritos se hace distinción entre los términos información y conocimiento (Alavi and Leidner, 2001). Entendiendo aquí información, como una colección de factores básicos que se convierten en conocimiento cuando son relevantes y tienen contexto (Alavi and Leidner, 2001).

Aristóteles⁵ fue el primero en distinguir que el conocimiento constaba de dos categorías: explícito y tácito (Baumard and Wauchope, 1999). El conocimiento explícito representa el conocimiento que ha sido codificado, usualmente en forma de texto. Para este caso en particular, el conocimiento explícito se refiere a las anotaciones hechas en los cuadernos de laboratorios que se presentan como evidencia para soportar la toma de decisiones. Por otro lado, el conocimiento tácito

⁵ Aristotle. Wikipedia, the free encyclopedia.

representa el conocimiento que los individuos han acumulado con base en sus experiencias y es usado en la toma de decisiones. Este tipo de conocimiento basado en la experiencia, es adquirido a diario en cada éxito o fracaso que el personal del laboratorio obtenga en sus ensayos; además, este tipo de conocimiento puede ser transmitido a otros colegas con menos experiencia. Sin embargo, debido a que el conocimiento se basa en gran medida en la experiencia propia de cada investigador, tanto el conocimiento explícito como el conocimiento tácito en muchas ocasiones son difícilmente transferibles (Alavi and Leidner, 2001). Esto se refleja en la falta de homogeneidad tanto en la forma de hacer las anotaciones en los cuadernos de laboratorio, como en la manera de representar su conocimiento para explicarlo a los demás. Esta falta de homogeneidad de formas y maneras de representación del conocimiento suponen un claro inconveniente a la hora de reutilizar esa información (Sandars and Heller, 2006).

2.3.2 Importancia de un Lenguaje Común. Uno de los desafíos que enfrentan los investigadores es aquel debido a la gran cantidad de información generada (Washington and Lewis, 2008, Helfrich, 2004). Esta situación se puede traducir en una limitante debida a la poca facilidad para analizar e interrelacionar estos datos. Organizar, y compartir información en varios niveles es una prioridad en todo tipo de laboratorio (Alavi and Leidner, 2001). Este escenario, donde se producen muchos datos y se manejan múltiples fuentes de información, ha generado la necesidad de tener formatos y vocabularios comunes; compartidos por las diferentes comunidades de practica. Por ejemplo, con el fin de facilitar el intercambio de información relativa a experimentos de microarreglos se generó un formato estándar para compartir este tipo de información –Minimum Information About Microarray Experiment, MIAME– (Brazma et al., 2001). Sin embargo, el contar con un formato estandarizado es solo una parte del problema; se hizo entonces evidente la necesidad de contar con un vocabulario controlado –Microarray Gene Expression Data (MGED) ontology–

(Stoeckert Jr and Parkinson, 2003). De esta manera la comunidad de microarreglos aseguró la replicabilidad y el correcto entendimiento de los experimentos; además facilitó la generación de repositorios para compartir estos experimentos. Los vocabularios controlados facilitan el intercambio de información, proveen un lenguaje común para la descripción de actividades, procesos e insumos biológicos; algunos incluyen listas de sinónimos, acrónimos, y abreviaturas (Organization, 2005).

La generación de vocabularios controlados es una actividad costosa y demorada; toma un gran esfuerzo por parte de la comunidad, y demanda recursos que no siempre están disponibles (Good et al., 2006). Recientemente, se han iniciado investigaciones que muestran como el proceso de generación de vocabularios controlados puede facilitarse usando principios de las folksonomías (Angeletou, 2008, Spiteri, 2007). De igual manera se estudia activamente como un vocabulario controlado puede beneficiarse de una folksonomía, y como estas dos entidades se relacionan entre si (Spiteri, 2007, Angeletou, 2008).

- Folksonomías. Folksonomía, también conocidas como taxonomías populares, es un termino acuñado por Thomas Vander Wal (Vander-Wal, 2007); según Vander Wal las folksonomías se valen de entornos comunitarios para la anotación de recursos, e.g. enlaces de noticias, fotografías, canciones, artículos especializados, etc. (Adam, 2004). Esto fomenta la colaboración y el intercambio ágil de información entre usuarios que giran en torno a un propósito, meta o interés común (Zhang et al., 2008). El proceso de anotación, se hace usando tags; estos son palabras claves que describen el contenido de un documento, al mismo tiempo le aportan atributos a dicho documento para facilitar su organización, clasificación y posterior recuperación (Heymann and Garcia-Molina, 2008).

Establecer una folksonomía consiste en definir relaciones entre: los recursos, un conjunto de tags y los usuarios. Por ejemplo, la relación que se establece cuando un investigador usa palabras clave o una frase para describir el contenido de las paginas de su cuaderno de laboratorio. Las tres entidades: recursos, tags y usuarios, se pueden relacionar entre si de diferentes maneras. La figura 5 presenta una tabla y una grafica que ilustran posibles relaciones entre estas tres entidades⁶.

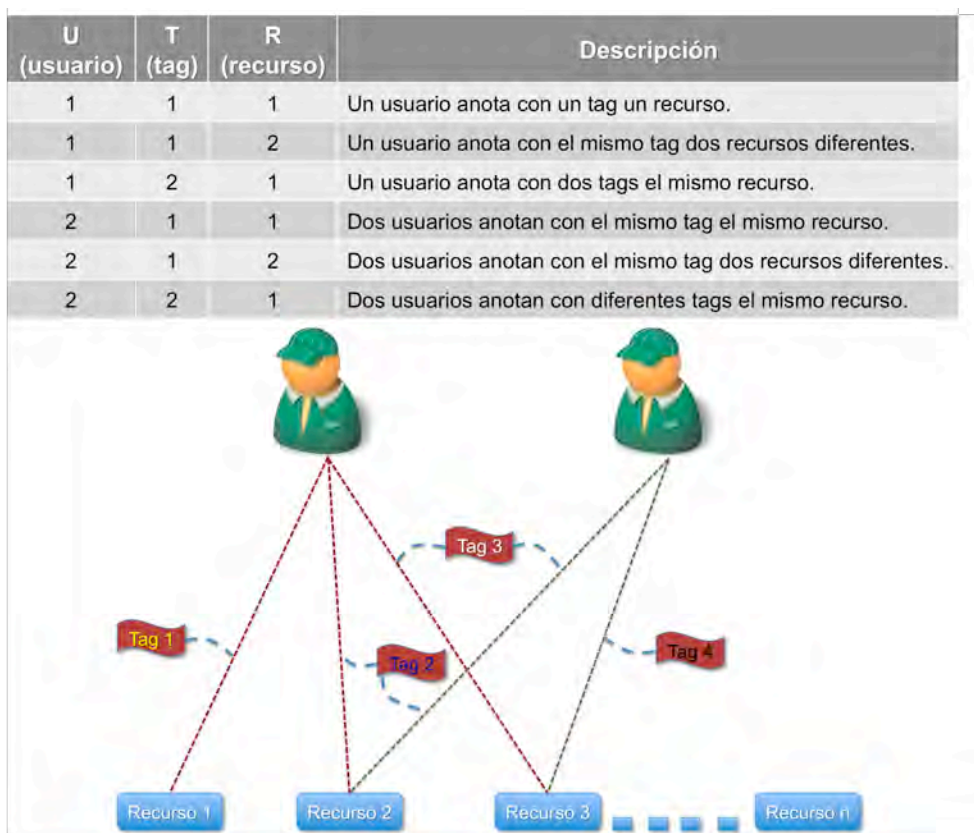


Figura 5. Posibles relaciones entre las entidades: usuarios, tags y recursos. Tomada del Proyecto Morfeo.

En síntesis, los términos “folksonomías”, “sistemas de anotación o tagueo social”, hacen referencia a una misma acción: organizar y clasificar información. Algunas

⁶ (2008) Integracion de folksonomias y ontologias, Proyecto Morfeo-MyMobileWeb.

aplicaciones web en las cuales los usuarios anotan toda clase de recursos se describen a continuación:

Delicious⁷: anteriormente conocido como “del.icio.us”. Es un servicio que permite agregar los marcadores de Internet, del inglés bookmarks; estos usualmente se manejan a través del navegador. Delicious facilita su manejo sobre una base de datos disponible vía Web; de ésta manera los recursos y anotaciones están disponibles independientemente del navegador. Delicious permite anotar los recursos favoritos de un usuario. Esta aplicación no solo almacena las anotaciones y las referencias a los objetos anotados, también permite compartirlos con otros usuarios de delicious y determinar cuantos de ellos comparten los mismos enlaces, ver figura 6.

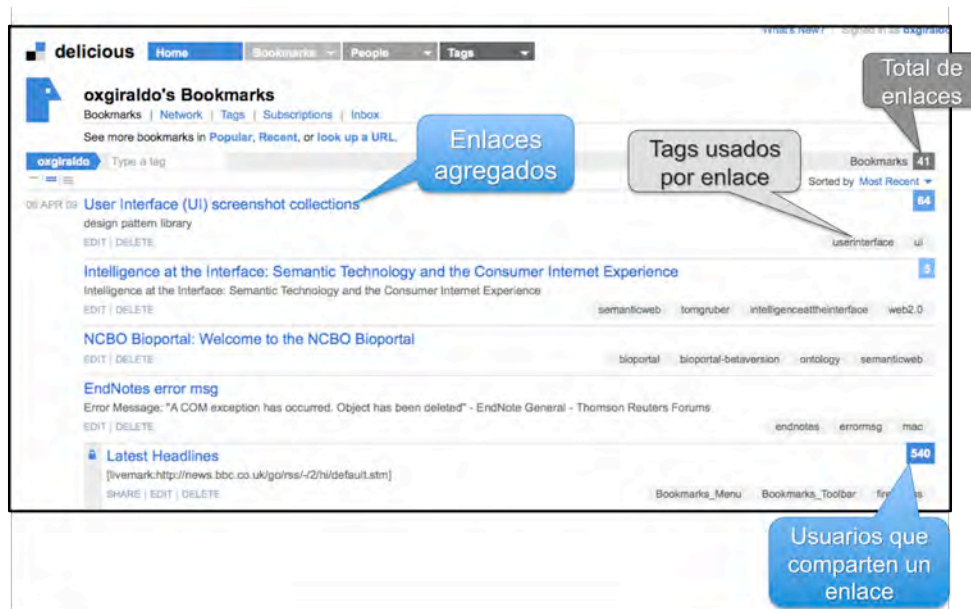


Figura 6. Captura de pantalla de delicious que muestra algunas de sus funciones: i) lista de enlaces, ii) tags usados por cada enlace y iii) número de usuarios que comparten un mismo enlace.

⁷ Delicious. (2003); Available from: <http://delicious.com/>.

Connotea⁸: este es un servicio cuyo principal objetivo es soportar la colaboración entre investigadores basada en conjuntos de anotaciones similares. Connotea maneja las anotaciones de manera equivalente a referencias bibliográficas. Permite también, guardar y compartir enlaces a sitios Web favoritos de manera similar a delicious, ver figura 7.

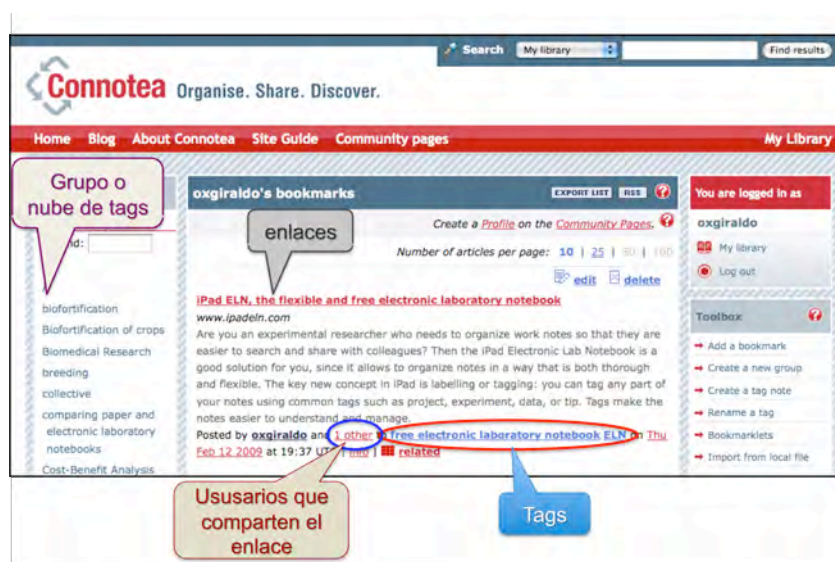


Figura 7. Captura de pantalla de connotea. Servicio en línea que permite guardar y compartir referencias bibliográficas y enlaces a sitios Web.

Picasa⁹: es un servicio de Google que permite almacenar y compartir fotos. Este recurso permite añadir tags a las fotos para facilitar su organización en álbumes. Actualmente Picasa “descubre” las caras de personas que aparecen en las fotos para facilitar al usuario, si así lo desea, ponerle nombre a las personas que aparecen en las fotos. Esto también le permite al usuario, organizar su colección de fotos en función de quien aparezca en ellas, ver figura 8.

⁸ Connotea. 2004; Available from: <http://www.connotea.org/>.

⁹ Picasa. 2006; Available from: <http://picasaweb.google.com>.

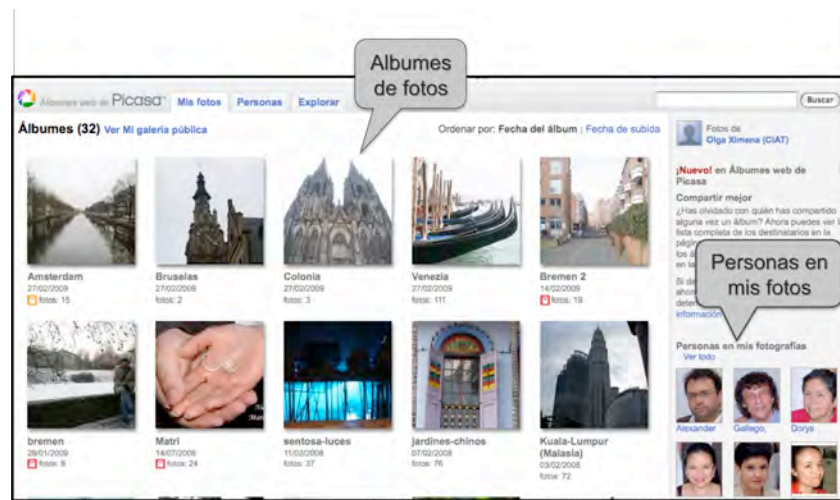


Figura 8. Captura de pantalla de Picasa. La aplicación permite subir, organizar y compartir fotos.

Facebook¹⁰: es una red social cuyo propósito es facilitar las relaciones a través del contenido generado por el usuario y los servicios que la misma red ofrece tales como: fotos, grupos, paginas, regalos, juegos, etc. Al igual que Picasa, los usuarios pueden etiquetar sus fotos y las personas que aparecen en ellas, ver figura 9.

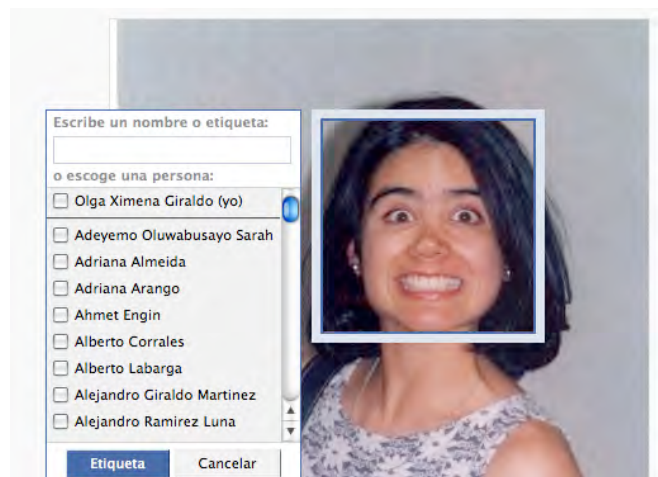


Figura 9. Captura de pantalla de facebook. Se ilustra la manera de agregar tags a las personas que aparecen en una fotografía.

¹⁰ ZUCKERBERG, M. (2004). Facebook; Available from: <http://www.facebook.com>.

Los sistemas de anotación o tagueo colaborativo antes descritos, permiten asociar uno o mas tags a un recurso digital (paginas web, fotografías, artículos científicos, etc). De igual manera facilitan la generación de tags personalizados que en principio le sirven al autor del tag y, una vez creados, generalmente le sirven a la comunidad (Heymann and Garcia-Molina, 2008). La búsqueda de recursos a partir de los tags disponibles, la facilidad para establecer asociaciones entre usuarios con similares recursos anotados con los mismos tags y la posibilidad de establecer asociaciones de parámetros como popularidad y calidad del recurso (Adam, 2004), son otras características comunes a los sistemas de anotación aquí descritos.

Las folksonomías presentan problemas tales como: ambigüedad, sinonimia y polisemia (Spiteri, 2007). Estos inconvenientes tienen que ver con la naturaleza abierta y libre del tagueo colaborativo; a su vez dependen de las habilidades, experiencias, intereses y estado mental de quien taguea (Nauman et al., 2008). Estos problemas generan imprecisión en el uso de términos lo cual conlleva a una mayor dificultad en la recuperación de recursos. Por ejemplo, los términos usados por investigadores para referirse a la Reacción en Cadena de la Polimerasa son: “PCR”, “prueba de amplificación”, “amplificación de...” , ver figura 10.

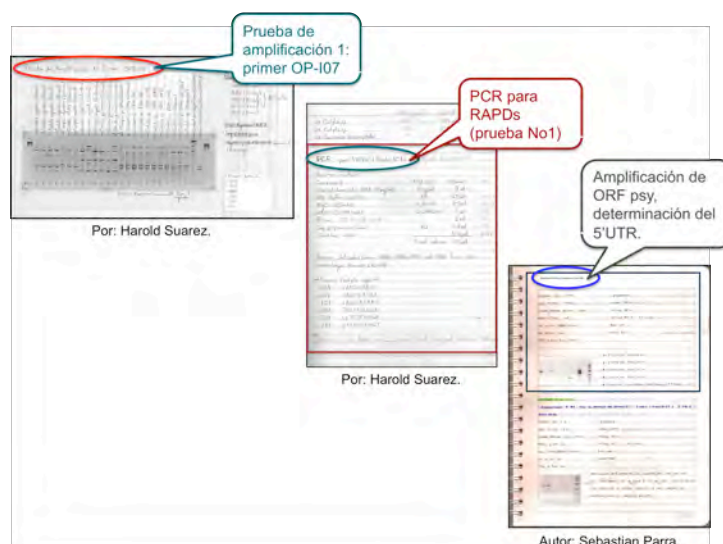


Figura 10. Ejemplos de la manera como investigadores anotan el procedimiento de PCR, con base al contenido del recurso y la experiencia.

El término “PCR” se refiere a las siglas que en inglés representan la frase Polymerase Chain Reaction (Reacción en Cadena de la Polimerasa, en español); este término es considerado un sinónimo del procedimiento experimental al cual hacemos referencia. Los términos “prueba de amplificación” y “amplificación de...” son imprecisos para referirse a la PCR. Si bien este procedimiento experimental es ampliamente usado para sintetizar o amplificar fragmentos específicos tanto de ADN como de ARN, no es la única técnica usada para la amplificación de ácidos nucleicos, existen otros métodos como la Ligase Chain Reaction (LCR) (Wiedmann et al., 2009) y la Self-Sustained Sequence Replication (3SR) (Mueller et al., 1997). Revisando la manera de anotar los procedimientos experimentales por parte de los investigadores, como en este caso se hizo con la técnica de PCR (figura 11), se pretende mostrar que a partir de un entorno colaborativo de usuarios con intereses y experiencias comunes, es posible resolver los problemas de ambigüedad en el léxico. Son los mismos usuarios expertos de un dominio de conocimiento quienes podrán generar, a partir de las folksonomías, vocabularios controlados. Estos vocabularios controlados pueden contribuir al mejoramiento de las bio-ontologías existentes¹¹.

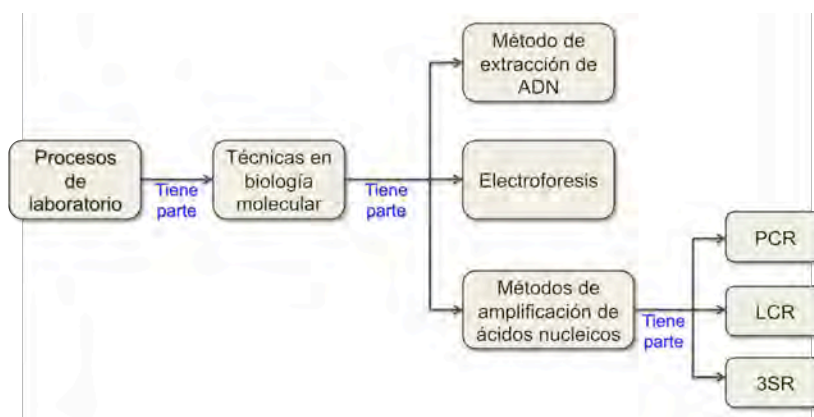


Figura 11. Mapa conceptual que ilustra una ontología parcial de la técnica de PCR. Tomada de OBI (Ontology for Biomedical Investigation) (Courtot et al., 2008) y Cancer Research and Management ACGT Master Ontology (Anguita et al., 2007).

¹¹ NCBO Bioportal. Available from: <http://stage.bioontology.org/>

- Vocabularios Controlados y Ontologías.

Vocabularios controlados. Los vocabularios controlados son generados para reducir o eliminar la redundancia o sinonimia en los términos¹². La definición de vocabularios controlados en el entorno digital, mejora la efectividad en la recuperación de la información (Angeletou, 2008).

Los vocabularios controlados son de cuatro tipos: las listas, los anillos de sinónimos, las taxonomías y los tesauros. La diferencia entre los cuatro tipos de vocabularios controlados esta determinada por la menor o mayor complejidad estructural que presentan. Por un lado, los listados y los anillos de sinónimos se limitan a incorporar la relación de equivalencia; por el otro, las taxonomías incorporan relaciones de jerarquía y los tesauros incorporan relaciones de equivalencia, jerarquía y asociativas¹².

Los vocabularios controlados pueden presentarse en el curso del desarrollo de ontologías. La figura 12 ilustra como los vocabularios controlados ser parte de una ontología. Independientemente de la metodología para desarrollar las ontologías, los vocabularios controlados al ganar mayor formalidad constituyen resultados parciales en el desarrollo de una ontología (Garcia-Castro, 2007).

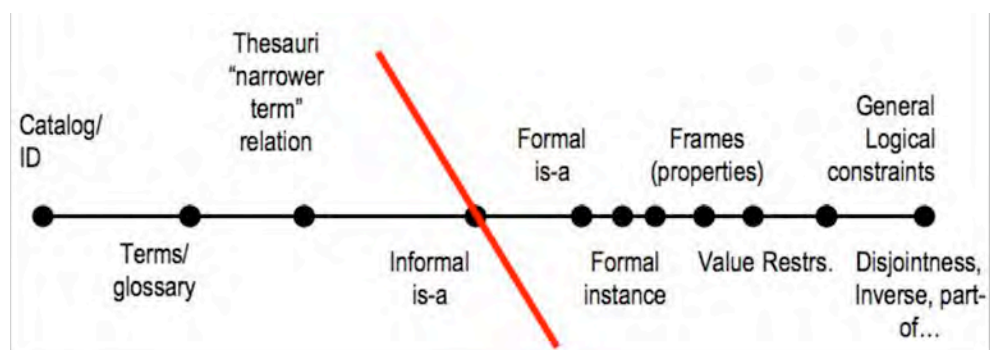


Figura 12. Transición de glosarios a ontologías. Tomado de (Garcia-Castro, 2007).

¹² (2008) Integración de folksonomías y ontologías, Proyecto Morfeo-MyMobileWeb.

Las ontologías en principio facilitan inferencias lógicas, los vocabularios controlados no; en ciencias biológicas estas han sido mayoritariamente usadas como vocabularios controlados (Smith et al., 2003). Un ejemplo de ello son la ontología de Genes, Gene Ontology (GO) (Ashburner et al., 2000), y la de microarreglos, Microarray Gene Expression Data (MGED) y Microarray ontology (MO) (Stoeckert Jr and Parkinson, 2003); estas se usan principalmente como vocabularios controlados para anotar y unificar datos a través de bases de datos biológicos existentes (García-Castro, 2007). Para entender un poco más el fundamento de los vocabularios controlados, en la figura 13, se ilustra su estructura conceptual.



Figura 13. Mapa conceptual que ilustra el fundamento de los vocabularios controlados.

Ontologías. La Web semántica es un área pujante nacida de la confluencia de la inteligencia artificial y las tecnologías Web; propone y requiere de nuevas técnicas y paradigmas para la representación del conocimiento para facilitar la localización, intercambio e integración de recursos a través de la Web (Zhang et al., 2008). Estas nuevas técnicas se basan en la introducción de conocimiento semántico explícito que describa y/o structure la información y servicios disponibles, de forma susceptible de ser procesada automáticamente por software (Seoane, 2007). Uno de los ejes principales de esta visión es la noción de ontología como herramienta clave

para alcanzar un entendimiento entre las partes (expertos de dominio, desarrolladores, programas) que participan de este conocimiento común (Socorro et al., 2008).

El termino “ontología” se deriva del griego (on= existencia, logos= razón) (Garcia-Castro, 2007). Una definición de ontología, la identifica con la “rama de la metafísica que estudia la naturaleza de la existencia” (Socorro et al., 2008). En el área de ciencias de la computación, las ontologías no son consideradas como una entidad natural que se descubre, sino como un recurso artificial que se crea (Mahesh, 1996). Una ontología podría comprenderse como, un entendimiento común y compartido dentro de un dominio de conocimiento (Socorro et al., 2008). Esta particularidad de las ontologías, de poder compartirse y reutilizarse en aplicaciones diferentes, explica en parte el gran interés suscitado en los últimos años en relación a la creación e integración de ontologías (Gangemi et al., 1998, Steve et al., 1998).

Según la definición de Gruber (Gruber, 1993), una ontología constituye “una especificación formal y explícita de una conceptualización”. En esta definición, “conceptualización”, se refiere a un modelo abstracto de algún fenómeno del mundo del que se identifican los conceptos que son relevantes; “explícito”, hace referencia a la necesidad de especificar de manera conciente los distintos conceptos que conforman una ontología; “formal”, indica que la especificación debe representarse por medio de un lenguaje de representación formalizado (Garcia-Castro, 2007).

Por otro lado, Weigand (Weigand and Hoppenbrouwers., 1997) define ontología como una base de datos donde se describen conceptos del mundo o de algún dominio del conocimiento específico, sus propiedades y como se relacionan los conceptos entre sí.

De acuerdo a las definiciones antes descritas sobre el termino “ontología”, en sentido general, una ontología es la base del contenido semántico. El contenido semántico, por ejemplo, en ciencias de la información, no es su “significado”. La palabra "ADN", por ejemplo, tiene un significado concreto, que se puede buscar en el diccionario.

Sin embargo, si a través del uso de un computador, se introduce la palabra "ADN" (ácido desoxirribonucleico), para el sistema no es más que una cadena de caracteres. Por lo tanto, para darle significado, se deberá codificar el concepto de una manera que el computador pueda "entenderlo", o por lo menos ser capaz de gestionar el concepto como tal.

Una manera de agregar contenido semántico a una colección de elementos, es aplicando la teoría de conjuntos (Norris, 2008); es decir, se puede definir "ADN" como un tipo de biomacromolécula que hace parte del conjunto de los ácidos nucleicos. El computador puede manejar los conjuntos, debido a que son constructos matemáticos bien definidos, y así gestionar el concepto. Para el computador, son los conjuntos lo que importa, y la comunidad añade las etiquetas "ADN" y "ácido nucleico" para que esos conjuntos sean entendibles para los humanos, ver figura 14.

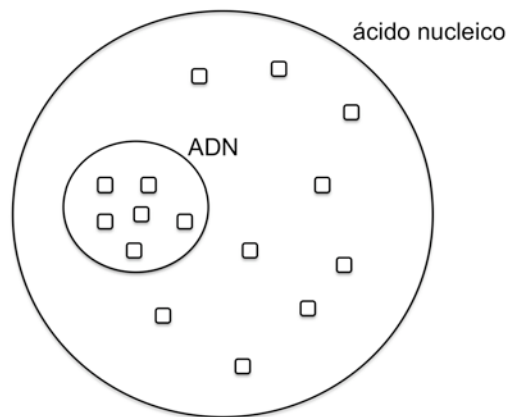


Figura 14. Modelo que representa la base de una ontología. Imagen tomada de (Aranguren et al., 2008).

El tipo de modelo que se acaba de describir es llamado ontología (Norris, 2008). En ese modelo, la clase "ADN" es una subclase o subconjunto de la clase "ácido nucleico", esto significa que todas las biomacromoléculas de ADN son ácidos nucleicos, pero no todos los ácidos nucleicos son ADN. Es importante subrayar que

los nombres de las clases son completamente triviales para el computador: esas mismas clases se podrían llamar "X" e "Y" y serían semánticamente equivalentes, lo que importa es la estructura (Aranguren et al., 2008).

Sin embargo las ontologías son mucho más complejas, que el ejemplo del ADN y el ácido nucleico. Según Gruber (Gruber, 1995), las ontologías se componen de:

conceptos: son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etc.

relaciones: representan la interacción y enlace entre los conceptos de un dominio. Suelen formar la taxonomía del dominio. Por ejemplo: es_un, parte_de, participa_en, etc.

funciones: son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología. Por ejemplo, pueden aparecer funciones como: función_de_enfriar, función_mecánica, etc.

instancias: se utilizan para representar objetos determinados de un concepto.

reglas de restricción o axiomas: son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Por ejemplo: "Si A y B son de la clase C, entonces A no es subclase de B". Los axiomas, junto con la herencia de conceptos, permiten inferir conocimiento que no esté indicado explícitamente en la taxonomía de conceptos. Algunos de los componentes antes descritos, se ilustran en la figura 15.

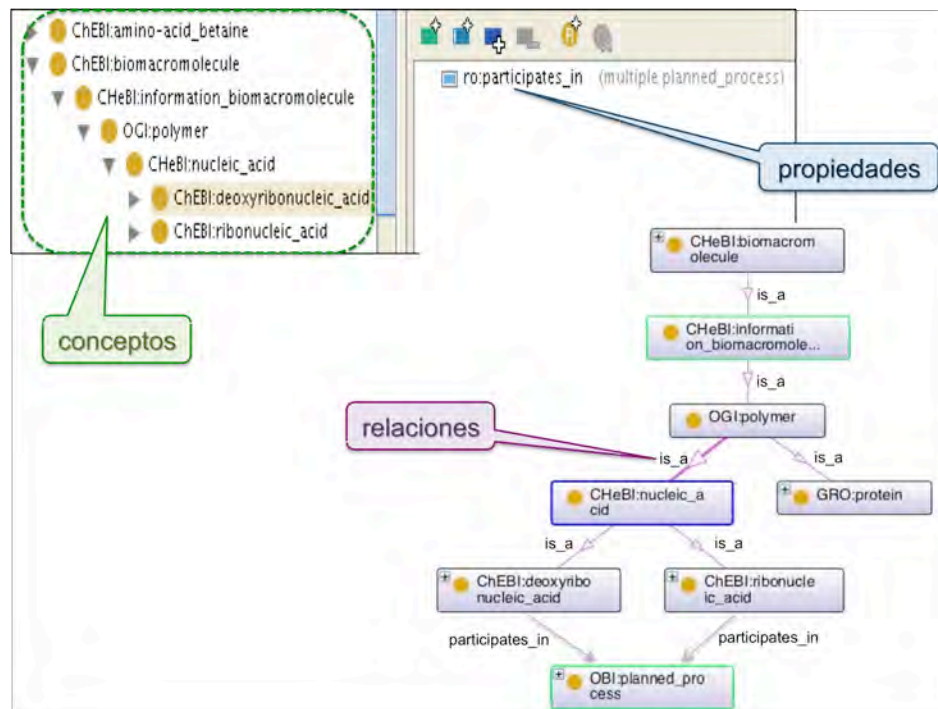


Figura 15. Se presenta un segmento de la ontología M4L, para ilustrar como protege-2000 (editor de ontologías), presenta los componentes de una ontología.

También existen distintos tipos de ontologías atendiendo diversos aspectos. Se pueden destacar las siguientes clasificaciones según el ámbito del conocimiento al que se apliquen:

Ontología de dominio (domain ontology o domain-specific ontology): modela un dominio específico, o parte del mundo (Garcia et al., 2010). Este tipo de ontología, representa los significados de los términos aplicados al dominio en cuestión. Un ejemplo de este tipo de ontologías, son aquellas que hacen parte del proyecto OBO (Open Biological and Biomedical Ontologies) (Smith et al., 2007). OBO contiene ontologías de dominio biomédico o biológico, que representan conocimiento a nivel molecular como PROteins (Natale et al., 2007) y CHEBI (Degtyarenko et al., 2008), hasta ontologías anatómicas a diferente nivel taxonómico. Por ejemplo en humanos existen ontologías como la Foundational Model of Anatomy (FMA) (Rosse and Mejino, 2003) y Edinburgh Human

Developmental Anatomy (EHDA) (Hunter et al., 2003). A nivel de plantas esta Plant Ontology (PO) (Mizoguchi et al., 2000), ver figura 16.

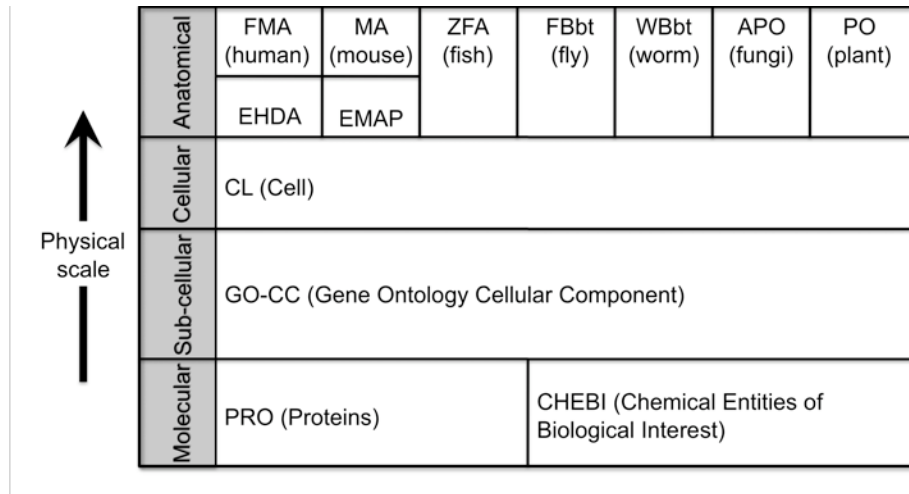


Figura 16. Ontologías pertenecientes al proyecto OBO, que representan el dominio biológico o biomédico desde nivel molecular hasta organismos completos pertenecientes a diferentes taxones. Tomado de [76]

Ontología superior o general (upper ontology, foundation ontology o Top-level ontology): estas ontologías describen conceptos generales sobre el espacio, tiempo o eventos que son independientes de un dominio específico [69]. Estas ontologías genéricas son reutilizadas por diferentes dominios. Algunos ejemplos de ontologías superiores son: Dublin Core (DC) (Weibel, 2000), General Formal Ontology (GFO) (Herre et al., 2006), OpenCyc/ResearchCyc (Matuszek et al., 2006), Suggested Upper Merged Ontology (SUMO) (Pease and Niles, 2002), Descriptive Ontology for Linguistic and Cognitive Engineering DOLCE (Masolo et al., 2003) y Basic Formal Ontology (BFO) (Grenon et al., 2004). BFO, sirve de apoyo para la construcción de las ontologías de dominio, desarrolladas en áreas de investigación científica, por ejemplo, aquellas pertenecientes al proyecto OBO, ver figura 17.



Figura 17. Organización jerárquica de las ontologías generales y de dominio, con base al conocimiento que los dos tipos de ontología describe.

Dado que las ontologías son buenas candidatas para añadir semántica a la Web, y así facilitar el recuperar información útil en cualquier dominio del conocimiento, se recomienda a la comunidad científica tener en cuenta a la hora de diseñar una ontología recomendaciones como (Castro et al., 2006, Schober et al., 2009):

Claridad: una ontología debe poder comunicar de manera efectiva el significado de sus términos. Las definiciones deberán ser lo más objetivas posibles y las mismas, deben explicarse en lenguaje natural.

Coherencia: una ontología debe permitir hacer inferencias que sean consistentes con las definiciones.

Extensibilidad: deben anticiparse nuevos usos para así poder permitir extensiones y especializaciones.

Especificidad: se debe especificar a nivel de conocimiento, sin que dependa de una codificación particular a nivel de símbolo.

Precisión: debe hacerse la menor cantidad de "suposiciones" acerca del mundo modelado.

En síntesis, existe un grado de complejidad en el proceso de diseño y desarrollo de ontologías, pero vale la pena el esfuerzo ya que las ontologías resultan idóneas para aportar el debido formalismo a una aplicación ¹³.

Debido a la naturaleza formal de las ontologías se llegó a plantear que estas habían quedado, en cierta medida, obsoletas (Shirky, 2005) frente a las folksonomías; mismas que por su naturaleza informal, han tenido una irrupción exitosa en la manera de anotar recursos en la Web (Vander-Wal, 2007, Adam, 2004, Gruber, 2007). La realidad es más compleja, si bien en algunos dominios parece que se adaptan mejor las ontologías (por ejemplo: en el ámbito de las ciencias de la salud) y en otros se adaptan mejor las folksonomías (por ejemplo: portales comunitarios de compartición de medios audiovisuales), en esta tesis se ilustrará como ontologías y folksonomías pueden coexistir. Se pretende aprovechar la flexibilidad que ofrecen las folksonomías para proveer un mecanismo sencillo de anotación en el contexto de los cuadernos de laboratorio; en donde sea posible se estructurará la anotación sobre ontologías existentes –sin exponer la complejidad del modelo a los usuarios-. Además se espera que el prototipo en sí mismo llegue a ser un modelo de transición para el diseño de un cuaderno de laboratorio electrónico que permita la captura de información en formato digital de manera fácil, eficiente, y segura.

2.4 BIBLIOGRAFÍA

ADAM, M. (2004) Folksonomies - Cooperative Classification and Communication Through Shared Metadata. <http://www.adammathes.com>.

ALAVI, M. & LEIDNER, D. E. (2001) *Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research*

¹³ (2008) Integración de folksonomías y ontologías, Proyecto Morfeo-MyMobileWeb.

Issues, Management Information Systems Research Center, University of Minnesota.

- ANGELETOU, S. (2008) Semantic Enrichment of Folksonomy Tagspaces. *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg.
- ANGUITA, A., CALLE, G., SILVA, A., MAOJO, V., WEILER, G., BROCHHAUSEN, M., KUMAR, A., DURAND, P., KONDYLAkis, H. & DOERR, M. (2007) Consolidated Requirements on ontological approaches for integration of multi-level biomedical information.
- ARANGUREN, M. E., ANTEZANA, E., KUIPER, M. & STEVENS, R. (2008) Ontology Design Patterns for bio-ontologies: a case study on the Cell Cycle Ontology. *BMC Bioinformatics*, 9, -.
- ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D. P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M. & SHERLOCK, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25, 25-9.
- BAUMARD, P. & WAUCHOPE, S. (1999) ***Tacit knowledge in organizations***, London, SAGE.
- BRAZMA, A., HINGAMP, P., QUACKENBUSH, J., SHERLOCK, G., SPELLMAN, P., STOECKERT, C., AACH, J., ANSORGE, W., BALL, C. A., CAUSTON, H. C., GAASTERLAND, T., GLENISSON, P., HOLSTEGE, F. C. P., KIM, I. F., MARKOWITZ, V., MATESE, J. C., PARKINSON, H., ROBINSON, A., SARKANS, U., SCHULZE-KREMER, S., STEWART, J., TAYLOR, R., VILO, J. & VINGRONS, M. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *nature genetics* 29, 365-371.
- BUTLER, D. (2005) Electronic notebooks: A new leaf. *Nature*, 436, 20-21.
- CASTRO, A. G., ROCCA-SERRA, P., STEVENS, R., TAYLOR, C., NASHAR, K., RAGAN, M. A. & SANSONE, S. A. (2006) The use of concept maps during knowledge elicitation in ontology development processes - the nutrigenomics use case. *BMC Bioinformatics*, 7, -.
- COURTOT, M., BUG, W., GIBSON, F., LISTER, A. L., MALONE, J., SCHOBER, D., RYAN BRINKMAN, R. & RUTTENBERG, A. (2008) The OWL of Biomedical Investigations. *OWLED* 12.
- DEGTYARENKO, K., DE MATOS, P., ENNIS, M., HASTINGS, J., ZBINDEN, M., MCNAUGHT, A., ALCANTARA, R., DARSOW, M., GUEDJ, M. &

- ASHBURNER, M. (2008) ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Research*, 36, D344-D350.
- DELLAPORTA, S. L., WOOD, J. & HICKS, J. B. (1983) A plant DNA miniprep: Version II *Plant Molecular Biology Reporter*, 1, 19-21.
- ELLIOTT, M. (2006) Electronic Study Management. New Tools for Improving the Efficiency of Pre-Clinical R&D. *Scientific Computing*, 7.
- GANGEMI, A., PISANELLI, D. M. & STEVE, G. (1998) Ontology Integration: Experiences with Medical Terminologies. *N. Guarino*, 163-178.
- GARCIA, A., O'NEILL, K., GARCIA, L. J., LORD, P., STEVENS, R., CORCHO, O. & GIBSON, F. (2010) Developing Ontologies within Decentralised Settings. IN CHEN, H. (Ed.) *Semantic e-Science, In Press*. Springer Science+Business Media.
- GARCIA-CASTRO, A. (2007) DEVELOPING ONTOLOGIES IN THE BIOLOGICAL DOMAIN *Institute for Molecular Bioscience*. Queensland, University of Queensland.
- GOOD, B., TRANFIELD, E. M., TAN, P. C., SHEHATA, M., SINGHERA, G. K., GOSSELINK, J., OKON, E. B. & WILKINSON, M. (2006) Fast, Cheap, and Out of Control: A Zero Curation Model for Ontology Development. *Pacific Symposium on Biocomputing*.
- GRENON, P., SMITH, B. & GOLDBERG, L. (2004) Biodynamic ontology: Applying BFO in the biomedical domain. *Ontologies in Medicine*, 102, 20-38.
- GRUBER, T. (1993) A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2), 199-220.
- GRUBER, T. (2007) Collective Knowledge Systems. *Journal of Web Semantics*.
- GRUBER, T. R. (1995) Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43, 907-928.
- HAQUIN, S., OEUILLET, E., PAJON, A., HARRIS, M., JONES, A., TILBEURGH, H., MARKLEY, J., ZOLNAI, Z. & POUPON, A. (2008) *Data management in structural genomics: an overview.*, Humana Press.
- HELFRICH, J. P. (2004) Knowledge management systems: coping with the data explosion. *DDT: BIOSILICO*, 2(1).

- HERRE, H., HELLER, B., BUREK, P., HOEHNDORF, R., LOEBE, F. & MICHALEK, H. (2006) General Formal Ontology (GFO), Part I: Basic Principles, Version 1.0. Leipzig, Institute of Medical Informatics, Statistics and Epidemiology (IMISE).
- HEYMANN, P. & GARCIA-MOLINA, H. (2008) Can Tagging Organize Human Knowledge?, Infolab Technical Report. CA, USA, Stanford University.
- HILLIARD, M. S., LARSON, D. L. & ROSENBERG, M. J. (2001) LIMS: A Suite of Database Tools for Laboratory Organization. *Computer-Based Medical Systems, 2001. CBMS 2001. Proceedings. 14th IEEE Symposium on.* Bethesda, MD, USA, <http://ieeexplore.ieee.org>.
- HUNTER, A., KAUFMAN, M. H., MCKAY, A., BALDOCK, R., SIMMEN, M. W. & BARD, J. B. L. (2003) An ontology of human developmental anatomy. *Journal of Anatomy*, 203, 347-355.
- JAYASHREE, B., REDDY, T., LEELADEVI, Y., CROUCH, J., MAHALAKSHMI, V., BUHARIWALLA, H., ESHWAR, K., MACE, E., FOLKSTERMA, R., SENTHILVEL, S., VARSHNEY, R., SEETHA, K., RAJALAKSHMI, R., PRASANTH, V., CHANDRA, S., SWARUPA, L., SRIKALYANI, P. & HOISINGTON, D. (2006) Laboratory Information Management Software for genotyping workflows: applications in high throughput crop genotyping. *BMC Bioinformatics*, 7, 383.
- KANARE, H. M. (1985) *Writing the Laboratory Notebook*, Washington D.C, American Chemical Society.
- KÖHL, K., BASLER, G., LÜDEMANN, A., SELBIG, J. & WALTHER, D. (2008) A plant resource and experiment management system based on the Golm Plant Database as a basic tool for omics research. *Plant Methods*, 4, 1-11.
- LYSAKOWSKI, R. (1997) Comparing Paper and Electronic Laboratory Notebooks. <http://www.censa.org>.
- MAHESH, K. (1996) Ontology Development for Machine Translation: Ideology and Methodology, MCCS-96-292. Computing Research Laboratory, New Mexico State University.
- MASOLO, C., BORGO, S., GANGEMI, A., GUARINO, N., OLTRAMARI, A. & SCHNEIDER, L. (2003) WonderWeb Deliverable D17: The WonderWeb Library of Foundational Ontologies, Preliminary Report Padova, National Research Council, Institute of Cognitive Sciences and Technology.
- MATUSZEK, C., CABRAL, J., WITBROCK, M. & DEOLIVEIRA, J. (2006) An Introduction to the Syntax and Content of Cyc. *AAAI Spring Symposium, 2006*.

- MIZOGUCHI, R., KOZAKI, K., SANO, T. & KITAMURA, Y. (2000) Construction and deployment of a plant ontology. *Knowledge Engineering and Knowledge Management, Proceedings*, 1937, 113-128.
- MUELLER, J. D., PÜTZ, B. & HÖFLER, H. (1997) Self-sustained sequence replication (3SR): an alternative to PCR. *Histochemistry and Cell Biology*, 431-437.
- MYERS, J., FOX-DOBBS, C., LAIRD, J., LE, D., REICH, D. & CURTZ, T. (1996) Electronic Laboratory Notebooks for Collaborative Research. *IEEE Fifth Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*. Stanford University, California, <http://collaboratory.emsl.pnl.gov>.
- NATALE, D. A., ARIGHI, C. N., BARKER, W. C., BLAKE, J., CHANG, T. C., HU, Z. Z., LIU, H. F., SMITH, B. & WU, C. H. (2007) Framework for a Protein Ontology. *BMC Bioinformatics*, 8, -.
- NAUMAN, M., KHAN, S., AMIN, M. & HUSSAIN, F. (2008) Resolving Lexical Ambiguities in Folksonomy Based Search Systems through Common Sense and Personalization. *SemSearch 2008*. CEUR-WS.org/Vol-334/.
- NORRIS, C. (2008) Some Versions of Platonism: Mathematics and Ontology According to Badiou. *Philosophical frontiers*, 3, 28.
- ORGANIZATION, N. I. S. (2005) ANSI/NISO Z39.19 - Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. *National Information Standards Organization*. NISO Press.
- PAJON, A., IONIDES, J., DIPROSE, J., FILLON, J., FOGH, R., ASHTON, A., BERMAN, H., BOUCHER, W., CYGLER, M., DELEURY, E., ESNOUF, R., JANNIN, J., KIM, R., KRIMM, I., LAWSON, C., OEUILLET, E., POUPON, A., RAYMOND, S., STEVENS, T., TILBEURGH, H., WESTBROOK, J., WOOD, P., ULRICH, E., VRANKEN, W., XUELI, L., LAUE, E., STUART, D. & HENRICK, K. (2005) Design of a Data Model for Developing Laboratory Information Management and Analysis Systems for Protein Production. *PROTEINS: Structure, Function, and Bioinformatics*, 58, 278-284.
- PASZKO, C., TURNER, E. & HINTON, M. (2002) *Laboratory Information Management Systems*, CRC Press.
- PEASE, A. & NILES, I. (2002) IEEE standard upper ontology: a progress report. *Knowledge Engineering Review*, 17, 65-70.
- POLONSKY, A., SIX ADRIEN., KOTELNIKOV MIKHAIL., POLONSKY VADIM., POLLY RENAUD. & PAUL., B. Semantic Laboratory Notebook. <http://knol.google.com>.

- REBBECK, J. (2005) What Makes a Great Science Lab Notebook? , <http://www.sciencebuddies.org>.
- RODRIGUEZ-YUNTA, M. (2008) EL CUADERNO DE LABORATORIO. Madrid, Universidad Complutense de Madrid.
- ROSSE, C. & MEJINO, J. L., JR. (2003) A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform*, 36, 478-500.
- SANDARS, J. & HELLER, R. (2006) Improving the implementation of evidence-based practice: a knowledge management perspective. *Journal of Evaluation in Clinical Practice*, 12, 341-346.
- SARINI, M., BLANZIERI, E., GIORGINI, P. & MOSER, C. (2004) From actions to suggestions: supporting the work of biologists through laboratory notebooks. *Proceedings of 6th International Conference on the Design of Cooperative Systems (COOP2004)*. French Riviera, France, IOSPress.
- SCHOBER, D., SMITH, B., LEWIS, S. E., KUSNIERCZYK, W., LOMAX, J., MUNGALL, C., TAYLOR, C. F., ROCCA-SERRA, P. & SANSONE, S. A. (2009) Survey-based naming conventions for use in OBO Foundry ontology development. *BMC Bioinformatics*, 10, -.
- SEOANE, C. (2007) La Web 2.0 y sus implicaciones en el mundo de la gestión de la información, Actas del VI Workshop CALSI. *Calsi*.
- SHIRKY, C. (2005) Ontology is Overrated: Links, Tags and Post-hoc Metadata. *EMERGING TECHNOLOGY CONFERENCE*. San Diego, CA, <http://conferences.oreillynet.com>.
- SMITH, B., ASHBURNER, M., ROSSE, C., BARD, J., BUG, W., CEUSTERS, W., GOLDBERG, L. J., EILBECK, K., IRELAND, A., MUNGALL, C. J., THE OBI CONSORTIUM, LEONTIS, N., ROCCA-SERRA, P., RUTTENBERG, A., SANSONE, S.-A., SCHEUERMANN, R. H., SHAH, N., WHETZEL, P. L. & LEWIS, S. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25(11), 1251 - 1255.
- SMITH, B., WILLIAMS, J. & SCHULZE-KREMER, S. (2003) The ontology of the gene ontology. *AMIA Annual Symposium Proceedings*
- SOCORRO, R., SIMÓN, A., VALDÉS, R., FERNÁNDEZ, F. O., ROSETE, A., MORENO, M., LEYVA, E. & PINA, J. Las ontologías en la representación del conocimiento. Centro de Estudios de Ingeniería de Sistemas (CEIS), Instituto Superior Politécnico "José Antonio Echeverría" (CUJAE).

- SPARROW, J. (1998) *Knowledge in Organizations: Access to Thinking at Work*, SAGE Publications.
- SPITERI, L. F. (2007) Structure and form of folksonomy tags: The road to the public library catalogue. *Webology*.
- STEVE, G., GANGEMI, A. & PISANELLI, D. M. (1998) Integrating Medical Terminologies with ONIONS Methodology.
- STOECKERT JR, C. J. & PARKINSON, H. (2003) The MGED ontology: a framework for describing functional genomics experiments. *Comparative and Functional Genomics*, 4, 127-132.
- VANDER-WAL, T. (2007) Folksonomy. vanderwal.net.
- VOELKERDING, K., DAMES, S. & DURTSCHI, J. (2009) Next-generation sequencing: from basic research to diagnostics. *Clinical Chemistry*, 55(4), 641-58.
- WASHINGTON, N. & LEWIS, S. (2008) Ontologies: Scientific Data Sharing Made Easy. *Scitable by nature education*.
- WEIBEL, S. (2000) The Dublin core metadata initiative - The Frankfurt focus and the year 2000. *Zeitschrift Fur Bibliothekswesen Und Bibliographie*, 47, 3-13.
- WEIGAND, H. & HOPPENBROUWERS. (1997) Experiences with a Multilingual Ontology-based Lexicon for News Filtering, the TREVI project. Infolab.
- WIEDMANN, M., WILSON, W., BARANY, F. & BATT, C. (2009) Ligase chain reaction (LCR)--overview and applications. *Genome Research*, 1994. 3, S51-S64.
- ZHANG, Z., CHEUNG, K.-H. & TOWNSEND, J. (2008) Bringing Web 2.0 to bioinformatics. *BRIEFINGS IN BIOINFORMATICS*, 10(1), 1-10.

3. MANEJO DE CONOCIMIENTO: ANALIZANDO LOS CUADERNOS DE LABORATORIO

3.1 INTRODUCCIÓN

La experiencia colectiva de un laboratorio se refleja en los cuadernos de laboratorio en los cuales los investigadores detallan sus labores. El conocimiento con base en la experiencia de los investigadores se consigna en estos documentos como texto, imágenes, tablas, etc. Siendo los cuadernos de laboratorio artefactos no estructurados de baja tecnología, la correcta indexación de la información en ellos consignada es difícil. Mas aún, la información en ellos almacenada es inconexa de aquellas fuentes alternas de información tales como paginas Web, hojas electrónicas, documentos PDF, reportes de equipos de laboratorio, etc. De igual manera la información está desarticulada de aquella almacenada en otros cuadernos de laboratorio; incluso en aquellos casos en los cuales un experimento se inicia en una pagina y su continuación solo se consigna varias paginas mas adelante en el mismo cuaderno.

Dificultades similares en relación al manejo de conocimiento almacenado en documentos no estructurados han sido reportadas también en el dominio de organizaciones comerciales (Paganelli et al., 2006). Los documentos, estructurados y no estructurados, proveen un recurso que describe la experiencia, el *know-how*, de las organizaciones (Paganelli et al., 2006). En algunos casos son los documentos su principal producto; por ejemplo, contratos, reportes de consultoría, encuestas a consumidores, etc. La organización y facilitación de recuperación de información en

repositorios de documentos es una necesidad creciente y en la cual se hacen inversiones altas debido a su importancia (Olsen, 2004).

Los cuadernos de laboratorio, son un tipo particular de documentos propios de actividades relativas a la investigación. Este tipo de documentos no son auto descriptivos y, además, son altamente idiosincráticos; su organización depende de cada investigador, lo cual dificulta la recuperación de la información. Esta dificultad se acentúa en la medida en que se generan mas cuadernos y se tienen mas medios alternos de información.

Una practica adoptada por los investigadores para identificar el conocimiento consignado en los cuadernos, es haciendo uso de “*anotaciones*”. Entiéndase por anotación un titulo, comentario, una nota, una explicación o cualquier texto que se pueda adjuntar a un documento con el propósito de describir su contenido. Idealmente los documentos tendrían una capa de descripciones embebidas; este tipo de documento inteligente ha sido definido por Uren et al (Uren et al., 2005), como documentos que “*conocen*” su propio contenido con el fin de que los procesos automatizados “*sepan que hacer con el*”. Estudiando las prácticas de anotación sobre recursos Web y aquellas que usan los investigadores en sus cuadernos de laboratorio, el presente estudio pretende facilitar el enriquecimiento de metadatos sobre este tipo de documentos.

Con este propósito se analizaron 408 paginas de 12 cuadernos de laboratorio del personal que labora en el proyecto de Agro Biodiversidad y Biotecnología del CIAT. Diez de esos cuadernos, contienen información relacionada con: Mapeo genético y análisis de QTLs, estudios de diversidad, transformación genética y cultivo *in Vitro* de tejidos. De igual manera, se estudiaron 2 cuadernos usados para llevar un registro sobre la extracción de la enzima Taq polimerasa, la cual es materia prima para la replicación *in Vitro* de ADN, proceso fundamental en la biología molecular de plantas.

El presente análisis se hace considerando algunas de las buenas practicas presentadas por Kanare (Kanare, 1985). La selección de los criterios analizados en este capitulo fue determinada por su relevancia en aspectos relativos al manejo de conocimiento. Para presentar el análisis de los cuadernos de laboratorio, este capitulo esta organizado así: en la sección 3.2, se analizan las características físicas de los cuadernos de laboratorio estudiados, en la sección 3.3, se describe el formato usado por el personal de laboratorio para llevar los cuadernos, en la sección 3.4, se presenta que tipo de datos encontrados en los cuadernos, en la sección 3.5, se expone la importancia del uso de metadatos para mejorar la recuperación de información consignada en los cuadernos y por ultimo se presentan los comentarios finales.

3.2 CARACTERÍSTICAS FÍSICAS DE CUADERNOS DE LABORATORIO ANALIZADOS

Los cuadernos revisados en este estudio, cumplen con los siguientes estándares:

- Tienen una encuadernación permanente, es decir, son de tapa dura, con hojas cosidas y enumeradas.
- El papel es resistente al rasgado y al derrame de ciertas sustancias en el laboratorio y es rayado lo cual resulta útil para la escritura y elaboración de todo tipo de esquemas.
- Los cuadernos, son además portables, y resistentes a caídas; esto facilita que el investigador pueda llevar el cuaderno a donde sea necesario.

En la figura 18, se ilustran las características físicas de los cuadernos de laboratorio analizados.

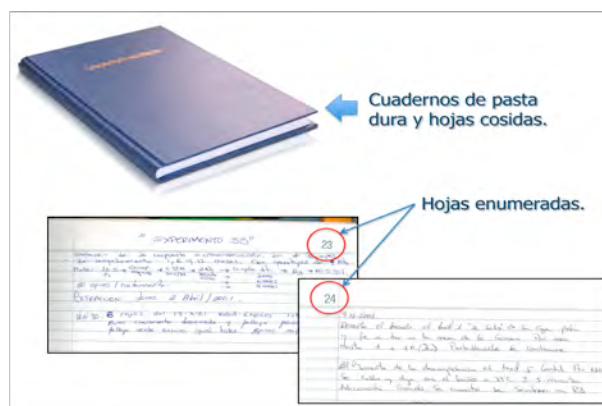


Figura 18. Características físicas de los cuadernos de laboratorio del proyecto Agro Biodiversidad y Biotecnología del CIAT.

3.3 FORMATO USADO EN LOS CUADERNOS DE LABORATORIO ANALIZADOS

A continuación se analizan las características principales que según Kanare (Kanare, 1985), deben ser consideradas al llevar un cuaderno de laboratorio: i) decir exactamente lo que se hizo y cuando, ii) dejar claro quien lo hizo, iii) ser permanente y verificable.

3.3.1 Decir Exactamente lo que se Hizo y Cuando. Para analizar este punto se planteó la pregunta: ¿está consignada la información suficiente para replicar un experimento?

Difícilmente alguien diferente al autor del experimento logra la reproducción del mismo. Esto debido a las observaciones que se presentan a continuación:

- Cuaderno de laboratorio como diario de trabajo. El cuaderno de laboratorio es concebido por los investigadores como un diario de trabajo personal, donde se consigna la información de cada actividad o experimento realizado principalmente de manera manuscrita y escrita para el entendimiento del autor del experimento. Para el investigador, el compartir la información no está contemplado directamente.

- Nivel de detalle. El nivel de detalle de la información consignada en los cuadernos es variable dependiendo de la experiencia de quien registra el experimento y/o de la frecuencia de realización de un determinado experimento. Por ejemplo, cuando se realiza un procedimiento por primera vez o su uso es poco frecuente (1 o 2 veces por año), el investigador en su cuaderno especifica el protocolo a seguir, hace una lista de los reactivos a usar, y las correspondientes cantidades utilizadas (volúmenes, pesos, concentraciones, etc.), además, escribe las observaciones durante el transcurso de la práctica, etc.; sin embargo, ese nivel de detalle se pierde cuando una práctica experimental es usada casi a diario, como la técnica de PCR o el proceso de visualización del producto amplificado, donde algunos investigadores escriben los reactivos a usar pero no sus concentraciones, tampoco es común encontrar las especificaciones del programa de PCR a usar para la amplificación de un determinado ADN, otros no registran la concentración del gel de agarosa o poliacrilamida usado para observar el producto amplificado, ni tampoco, el volumen servido del producto amplificado en un gel. El anterior ejemplo demuestra, que cuanto mayor es la experiencia de un investigador en el laboratorio y mayor la frecuencia de realización de una determinada actividad es menor el nivel de detalle de la información consignada en los cuadernos.

- Procesos asociados a diferentes proyectos consignados en un mismo cuaderno. En ocasiones los investigadores no pueden seguir de manera consecutiva y sin interrupciones los pasos que rigen un determinado proyecto de investigación debido a sus múltiples responsabilidades en el laboratorio. Para ilustrar esto se presenta el siguiente escenario:

Escenario

Un investigador a cargo de un proyecto relacionado con mapeo genético necesita realizar, a groso modo, los siguientes pasos:

1. Cosecha del material vegetal.
2. Extracción, cuantificación y dilución del ADN.
3. Proceso de estandarización y evaluación de un determinado tipo de marcador molecular, mediante el uso de la técnica de PCR.
4. Análisis de resultados.

Si además de la realización de dicho proyecto, el investigador debe cumplir con otras actividades dentro del laboratorio, en su cuaderno será imposible visualizar de manera consecutiva la ejecución de los 4 pasos presentados para llevar a cabo el proyecto de mapeo genético.

De acuerdo con el anterior escenario, fue común encontrar que un experimento iniciado en las paginas 1 y 2, continuara en las paginas 5, 6 y 7 y se finaliza en la pagina 20. Esta situación puede deberse a: i) que ciertos pasos de una determinada actividad experimental toman horas, días o meses; entre tanto otras actividades se habrán iniciado, generando el consecuente registro en el cuaderno de laboratorio, y ii) que ciertos pasos del procedimiento experimental requieran el

uso de un equipo no disponible en el momento, ya sea porque otro colega lo tenga en uso o por daño temporal del mismo.

También se analizó “cuando se hizo una determinada actividad o experimento”. El personal de laboratorio registra la fecha al inicio de cada actividad realizada, seguida por un título descriptivo de la experiencia llevada a cabo. Estos dos tipos de datos le facilitan al investigador la recuperación de la información a corto plazo; sin embargo, debido a las múltiples tareas, y el paso del tiempo, este tipo de registros pierde su utilidad. Una alternativa para mejorar la recuperación de la información es la elaboración de un índice o tabla de contenido; sin embargo siendo esta una labor manual que demanda tiempo, tales índices o tablas de contenido no se encontraron en los cuadernos analizados.

En resumen, algunos factores que afectan el manejo del conocimiento consignado en los cuadernos de laboratorio son:

- Falta de estándares: se hace difícil para cualquier persona el entender la información consignada en los cuadernos de laboratorio. Esto dificulta, por ejemplo, el reproducir experimentos.
- Falta de estrategias generales de indexación en los cuadernos de laboratorio: esto hace difícil la recuperación de información.
- Falta de estructura: dificulta el seguimiento de eventos no continuos registrados en los cuadernos.

3.3.2 Dejar Claro Quien lo Hizo. Algunos de los cuadernos analizados contienen una etiqueta en la tapa y/o el lomo del cuaderno con datos tales como: nombre, número del cuaderno y proyecto de investigación. Esa etiqueta, es la única manera para saber quien es el creador/autor de toda información consignada en el cuaderno porque las paginas no están marcadas con los datos del dueño.

Debido a que la etiqueta es un adhesivo, dicha etiqueta es vulnerable al desgaste o perdida. Se recomienda que al inicio de cada experiencia se incluya el nombre de la o las personas que la realizan; sin embargo el personal de laboratorio no lo hace (Kanare, 1985).

3.3.3 Ser Permanente y Verificable.

- Permanencia de la información. Se analizan aquí características relativas a la conservación del conocimiento en ellos consignado. La pasta del cuaderno, además de ser dura, es plastificada. Protege así las hojas del cuaderno de derrames de líquidos –siempre que el cuaderno permanezca cerrado. De igual manera, el papel usado es resistente al derrame de algunas sustancias; protegiendo así, también, la información consignada.

Una regla de oro al llevar un cuaderno de laboratorio es no arrancar hojas (Kanare, 1985). Una característica de los cuadernos analizados es el contar con páginas enumeradas. Esto facilitó el determinar que no faltaban hojas. Por el contrario, se evidenció que es frecuente añadir hojas; pudiendo ser éstas tablas, fotografías o cualquier tipo de ilustración proveniente de equipos de laboratorio, paginas Web, etc.

Los cuadernos de laboratorio analizados en su mayoría están escritos con tinta indeleble, pero en ocasiones los errores fueron tachados completamente o eliminados con corrector liquido. También se encontró información escrita a lápiz, sin manera de comprobar si se había borrado información. De cualquier modo, en el caso de un cuaderno de laboratorio no importa con que instrumento se esté escribiendo; lo importante es no borrar o tachar completamente (Kanare, 1985).

Igualmente se observó que cada investigador tiene en su escritorio la colección de cuadernos de laboratorio de su autoría. Además de aquellos cuadernos dejados por antiguos colegas, que trabajaban en proyectos afines, y que ya no hacen parte de la institución. En relación a estos cuadernos antiguos, aquellos mas recientes y los “actualmente en uso”, en el laboratorio no se lleva un registro que permita determinar “quien tiene los cuadernos dejados por quien” o “cuantos cuadernos han sido escritos por un investigador”.

En resumen, la durabilidad y vida útil de los cuadernos analizados depende principalmente de: i) las características físicas, ii) el no arrancar hojas y iii) uso de tinta indeleble para la escritura. Una recomendación, producto del estudio de los cuadernos, con relación a la permanencia de la información consignada es el hacer cumplimiento estricto de la regla no borrar o tachar completamente. Cuando se considere necesario, solo se deberá trazar una línea sobre la parte del escrito en cuestión. De esta manera la información estará disponible para otros investigadores (Kanare, 1985).

- Información verificable por el autor. Se evaluó también si la información era verificable. El termino “verificable” según el diccionario de la RAE, significa que permite comprobar su verdad y examinar el método por el que se ha alcanzado. De acuerdo con esa definición, retomamos la pregunta de la sección 3.3.1 (¿está consignada la información suficiente para replicar un experimento?), para analizar si el conocimiento consignado en los cuadernos evaluados podía ser verificable por el autor y por otros colegas. Como se describió en la sección 3.3.1, la información contenida en los cuadernos es en su mayoría manuscrita, de manera personalizada y el nivel de detalle depende de la experiencia y/o frecuencia con que se realice una determinada actividad. Los investigadores procuran mantener un registro histórico tanto de éxitos como de fracasos de sus practicas experimentales; al final de cada procedimiento pegan al cuaderno una

foto, tabla o cualquier tipo de ilustración que soporta los resultados obtenidos. De esta manera el autor de la información generalmente logra reproducir de manera exitosa un determinado procedimiento experimental.

- Información verificable por terceros. La falta de estructura, omisión de información y escritura ilegible son las principales razones por las cuales alguien con los conocimientos necesarios no pueda reproducir, comparar o compartir la información consignada en un cuaderno. Debido a este problema a continuación se presentan las recomendaciones que, según Kanare (Kanare, 1985), deben tenerse en cuenta al momento de consignar la información en un cuaderno de laboratorio:

1. La información consignada debe ser fiel a la realidad,
2. se deben escribir los resultados negativos, porque es la única forma de corregir los errores,
3. para la escritura se debe procurar usar un estilo impersonal, por ejemplo,
 - a. Se disuelve el producto A..... en lugar de: disolví el producto A...
 - b. Se procede a separar la mezcla... en lugar de: para separar la mezcla realicé tal procedimiento...
 - c. El rendimiento de la reacción es... en lugar de: me salió un rendimiento de...
4. Si el procedimiento experimental incluye el seguimiento de un protocolo, no es necesario que sea transcrito. Debe escribirse el nombre del protocolo a seguir y su correspondiente referencia.

5. Se deben indicar las medidas de seguridad requeridas en función de las características del experimento y de los reactivos.
6. Una vez comenzada la experiencia se deben escribir las cantidades realmente usadas (volúmenes, pesos, concentraciones), escribir cualquier modificación realizada con respecto al protocolo inicialmente previsto, los hechos observados durante el proceso: cambio de color, formación de precipitados, etc. y finalmente los rendimientos obtenidos (pesos, volúmenes, concentraciones o porcentajes).

3.4 TIPO DE DATOS ENCONTRADOS EN LOS CUADERNOS

Davenport y Prusak (Davenport and Prusak, 1998), definen “dato” como la mínima unidad semántica, o elementos primarios de información; siendo por sí solos irrelevantes como apoyo a la toma de decisiones. De forma general, los datos son la materia prima bruta. Al atribuírsele algún significado, pudiendo así relacionarse con otros datos, se convierten en información.

En los cuadernos de laboratorio se almacenan datos no estructurados, generados principalmente de manera manuscrita; no entendibles ni procesables por un computador. Con el propósito de determinar el tipo de contenido y datos disponibles en los cuadernos de laboratorio, fueron revisadas 408 páginas escogidas libremente por los mismos investigadores. En las páginas analizadas se identificaron 2421 datos. Estos datos fueron organizados dentro de tres grupos principales: 1) Información manuscrita, 2) información digital, e 3) información digital con anotaciones manuscritas. El número promedio de datos registrados por página fue de $5.93 \approx 6$. La labor de análisis se llevo a cabo de manera manual, lo cual fue posible debido al tener un conocimiento operativo relativo al llevar un

cuaderno de laboratorio. A continuación se describen los tres tipos de datos identificados en este estudio.

3.4.1. Información manuscrita. Como su nombre lo indica se refiere a todo tipo de anotación manuscrita creada y editada directamente en los cuadernos de laboratorio. Dentro de este grupo de datos se pudo identificar 10 subgrupos: 1) verificación de la actividad realizada, 2) palabras claves, 3) listados, 4) resultados parciales, 5) guías como se hace (how-to), 6) tareas por hacer (to-do), 7) fechas, 8) formulas algebraicas, 9) ruta de archivos electrónicos y 10) fuente de la información. El criterio para organizar la información manuscrita dentro de estos 10 subgrupos se presenta a continuación.

- Verificación de cada paso o actividad experimental. Con el propósito de no dejar nada a la memoria es costumbre del investigador, el verificar cada paso o actividad experimental realizada usando un visto bueno (✓). También se encontró el uso de palabras monosílabas como ya! u OK! para dejar claro la culminación de un paso (figura 19).

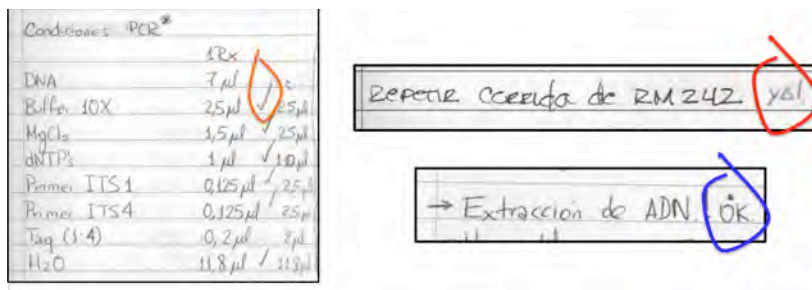


Figura 19. Símbolos usados por el investigador para verificar cada paso o cada actividad experimental realizada.

- Palabras claves. Quienes llevan un cuaderno de laboratorio, hacen uso de palabras claves que les ayuda a recuperar la información allí almacenada. El investigador emplea diferentes maneras para hacer notar las palabras claves del resto del texto manuscrito. Las mas comunes son: 1) uso de un marcador resaltador sobre una palabra o frase, 2) escritura con color diferente al resto del texto, 3) escritura en letras mayúsculas y 4) uso de palabras subrayadas (figura 20).

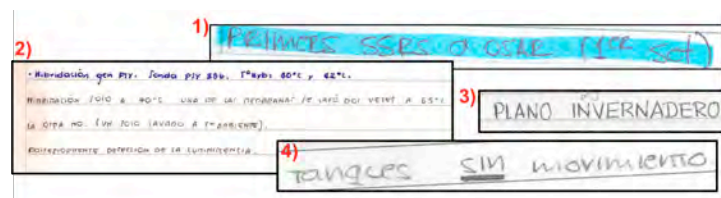


Figura 20. Uso de palabras clave.

- Listados. Se refiere a la enumeración de datos que son necesarios para realizar una actividad experimental. Los listados comúnmente registrados en los cuadernos analizados fueron de: A) cebadores, B) reactivos, D) material vegetal y D) cuantificación de ADN (figura 21).

A	TEMPERATURA	INDICADOR	INDICADOR
1. EM 11	55°C	Polim.	110-150pb
2. EM 14	55°C	H	110-150pb
3. EM 15	55°C	H	110-150pb
4. EM 16	55°C	H	110-150pb

B	STOCK	FINA	VOL
1. EM 11	1000 mM	100 mM	50 ml
2. EM 14	1000 mM	500 mM	250 ml
3. EM 15	1000 mM	100 mM	50 ml
4. EM 16	1000 mM	100 mM	50 ml

C	SE	INDICADOR	INDICADOR
1. 48-CT17182 (2)	1. Cebilla agreste	120	
2. 49-CT17183 (2)	2. Cebilla agreste	110	
3. 50-CT17184 (18)	3. Cebilla agreste	120	
4. 51-CT17185 (11)	4. Cebilla agreste	90	
	5. Cebilla agreste	80	

Figura 21. Tipo de listados encontrados en los cuadernos de laboratorio.

- Resultados parciales. Es la información conseguida después de llevar a cabo una actividad experimental. Este tipo de dato se presenta como párrafos cortos al final del procedimiento realizado y en ocasiones se usan íconos para describir los éxitos (☺) o fracasos (☹) (figura 22).

PER 436: 5 cajas del 23-X-00 Edad: 6 MESES- materna barto con poca senescencia tallos poco yemas verde, escuros buen follaje buen crecimiento y desarroll. Sin arte. Solo 2 cajas con 1C El 27-II-01.	
Nota: se corrió un gel de agar. con estas 2 encayas, y se obtuvo una amplif desigual de las muestras	
¡ SI HAY ACTIVIDAD! Los PCR amplificaron	☹ Resultado: Muy buena reproducción de las patrones de amplificación y buen polimorfismo interespecifico.

Figura 22. Forma de presentar los resultados de una actividad experimental.

- Guías como se hace (how-to). Se refiere a anotaciones que contienen instrucciones sobre como hacer ciertos procedimientos de laboratorio. El propósito de mantener estos “how-to”, es conseguir resultados reproducibles independiente de quien los haga. Anotaciones sobre las condiciones de una PCR, protocolos usados para extracción de ácidos nucleicos, o protocolos para realizar una electroforesis, son un buen ejemplo de guías “how-to” (figura 23).

B

EXTRACTION: DISCO DE ABRIL (QUINTERO, C)

- 1- Cortar el disco con una pastaseadora
- 2- Agregar 40 μ l de NaOH 0,25M
- 3- Calentar en baño de María a 100 $^{\circ}$ C x 30 seg
- 4- Agregar 40 μ l de HCl 0,25M
- 5- Agregar 20 μ l de TEH HCl 0,5M y pH 8.0
- 6- Calentar a 100 $^{\circ}$ C x 2 min
- 7- Hacer dilución a la mitad (20 μ l de muestra en 20 μ l de H₂O)

A

PCR para RAPDs (Prueba N°1) (15 Agosto de 2007)

Reaction mixture:

Component	Final conc.	Volumen
Diluted template DNA (5 μ g/ml)	1 ng/ μ l	5 μ l
10X Buffer reaction	1X	2.5 μ l
MgCl ₂ (25mM)	2.5 mM	2.5 μ l
dNTPs (5mM each)	0.2mM each	1 μ l
Primer (0.5 μ M)	6 μ M	1 μ l
Taq polymerase (10U)	1U	0.4 μ l
Distilled water		12.6 μ l
		Final volume: 25 μ l

Figura 23. Guías “how-to” encontradas en los cuadernos de laboratorio analizados.

- Tareas por hacer (to-do). Se refiere a las anotaciones hechas con relación a las actividades pendientes por hacer. Algunos ejemplos se pueden observar en la figura 24.

<p>NOVIEMBRE 12 DE 2008.</p> <p>a. Southern conda en gel de agarosa.</p> <p>COOPER LA NUESTRA EN EL GEL DE AGAROSA A 1% DURANTE TODA LA NOCHE</p>	<p>Referencia Corrida de RM 242</p>
<p>SELECCIONAR el cromosoma 1, 5, 10 de la población B3F2 LXB</p>	<p>!! Falta Mejorar condiciones de Amplificación !! → Verificar [J] primers y [J] final ADN.</p>
<p>→ Hacer una PCR-Touchdown bajando la temperatura desde 60 $^{\circ}$C a 55 $^{\circ}$C. Bajar la temperatura 1 $^{\circ}$C desde 60 $^{\circ}$C (por ciclo) hasta los 55 $^{\circ}$C. Luego continuar con la PCR a una temperatura de anillamiento de 55 $^{\circ}$C.</p>	<p>• para a seguir: Añalar LA BANDA DE 1776 pb. y CICHARLA</p>

Figura 24. Anotaciones relacionadas con tareas por hacer, encontradas en los cuadernos de laboratorio.

- Fechas. Se refiere al día, mes y año en que se realiza una actividad experimental consignada en el cuaderno. Este tipo de dato siempre se anota al principio de cada experimento y ayuda a facilitar la búsqueda y recuperación de la información consignada en los cuadernos (figura 25).

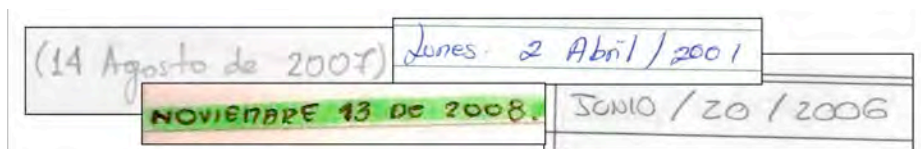


Figura 25. diferentes formas de anotar las fechas en los cuadernos de laboratorio.

- Formulas algebraicas. La formula matemática que presento mayor numero de registros en las paginas analizadas fue aquella usada para el cálculo de diluciones (concentración inicial x volumen inicial = concentración final x volumen final) (figura 26).

$$C_1 V_1 = C_2 V_2$$

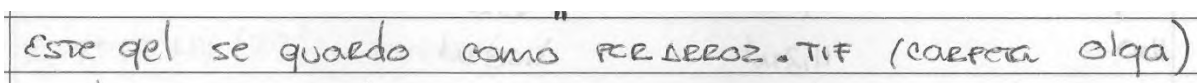
$$100 \text{ ng/ml} \times V_1 = 20 \text{ ng/ml} \times 100 \text{ ul}$$

$$V_1 = 20 \text{ ul ADN}$$

$$80 \text{ ul H}_2\text{O}$$

Figura 26. formula usada en el calculo de diluciones.

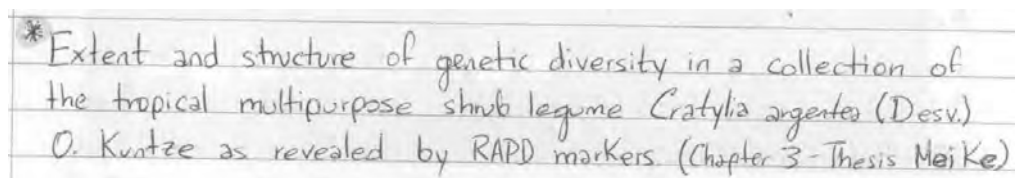
- Ruta de archivos electrónicos. Se evidenció el uso de medios alternos para el almacenamiento y manejo de información. Fue así como se encontraron hojas anexas a los cuadernos con tablas, fotografías, secuencias parciales de un gen, protocolos, etc. La proveniencia de estos medios alternos de información fue ocasionalmente especificada en los cuadernos (figura 27).



Este qel se guarda como REABROZ.TIF (carpeta olga)

Figura 27. anotación hecha acerca de una ruta de archivo electrónico.

- Fuente de la información. Es el lugar de donde se extrae información, por ejemplo monografías, artículos, patentes y todo tipo de contenedor de información (figura 28).



*Extent and structure of genetic diversity in a collection of the tropical multipurpose shrub legume *Cratylia argentea* (Desv.) O. Kuntze as revealed by RAPD markers (Chapter 3 - Thesis Mei Ke)

Figura 28. anotación relacionada con una fuente de información.

Los diez tipos de información manuscrita identificada en este estudio se resumen en la tabla 1. Estos datos fueron organizados de mayor a menor numero de ocurrencias en las 408 paginas analizadas. El total de datos manuscritos encontrados fue de 2252.

Tabla 1. Tipos de información manuscrita identificada en los cuaderno de laboratorio analizados.

		Información creada directamente en cuadernos de laboratorio	
Información manuscrita	verificación de cada paso o actividad realizada	uso de un visto bueno (✓) ya! OK!	1265 12 11
		Total	1288
	palabras claves	uso de marcador resaltador <u>subrayar palabras</u> escritura en otro color USO DE LETRAS MAYUSCULAS	292 76 57 9
		Total	434
	listados	material vegetal Vs. cebadores vectores de clonación material vegetal cuantificación de DNA cebadores reactivos	1 1 3 11 26 91
		Total	133
	resultados parciales	párrafos describiendo los resultados obtenidos uso de iconos para representar éxitos, e.j.: ☺ uso de iconos para representar fracasos, e.j.: ☹	46 32 24
		Total	102
	Como se hace (how-to)	condiciones de la PCR programas de la PCR protocolos de extracción de ácidos nucleicos protocolos de electroforesis en geles de acrilamida protocolos de electroforesis en geles de agarosa Protocolos de tinción de ácidos nucleicos	73 15 3 3 2 2
		Total	98
	tareas por hacer (to-do)	material vegetal por cosechar ácido nucleico por extraer ácido nucleico por cuantificar cebadores por estandarizar	24 21 17 15
		Total	77
	fechas	E.j.: dd/mm/aa	65
	formulas	C1xV1 = C2xV2	62
	Ruta de archivos electrónicos	E.j.: oxgiraldo/desktop/oxg/varios/cv-oxg	1
	Fuente de la información	E.j.: paginas web, artículo científico, tesis, etc.	1

2252

3.4.2. Información Digital. A este grupo pertenece aquella información que proviene de diferentes medios electrónicos la cual es impresa, recortada y pegada en los cuadernos de laboratorio. Dentro de este grupo de datos se pudo identificar 6 subgrupos: 1) fotografías, 2) listados, 3) resultados parciales, 4) guías how-to, 5) figuras, y 6) secuencias parciales. A continuación se presenta en detalle cada uno de estos 6 tipos de información digital.

- **Fotografías.** En los cuadernos de laboratorio se encontraron imágenes de electroforesis tanto en geles de agarosa como poliacrilamida obtenidas mediante una cámara fotográfica (figura 29).

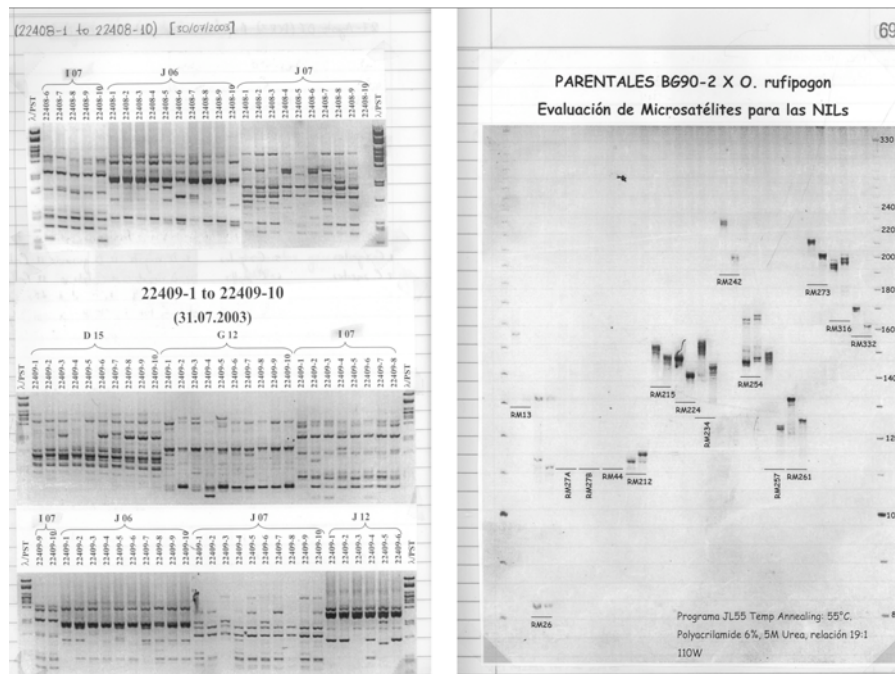


Figure 29. tipo de fotografías encontradas en los cuadernos de laboratorio.

- Listados. Los listados encontrados en formato electrónico se presentan a manera tabla. Algunos ejemplos son: a) listados del material vegetal, b) listados de cebadores, c) listados de evaluaciones fenotípicas, y d) listados de evaluaciones genotípicas con diferentes tipos de marcadores moleculares (figura 30).

A

Genotipos para Evaluación de SNPs de arroz Teos C.X. Glavito

Orden de numeración	Designación	Observaciones
1147	TOX 1858-102-484-B-01 1028-000-2021	del grupo de 533 genot.
21161	FL04062 (P-3) 2P-2P-M-04	del grupo de 533 genot.
1024	HO 10-2-002 17-ACC-01	del grupo de 533 genot.
41225	MAC 184-B-01 138-ACC-188	del grupo de 533 genot.
12144	CHYVICA 1-002 180-ACC-187	del grupo de 533 genot.
21223	CHYVICA 11-002 4-002 880-ACC-1951	del grupo de 533 genot.
12117	RNA YEMSAI 1-002 833-ACC-138	del grupo de 533 genot.
10164	ITC 4-002 80-1872 1872-ACC-1880	del grupo de 533 genot.
81587	AZUL-5NA-002 11-ACC-15	del grupo de 533 genot.
10162	CALCULO 83-102 859-ACC-18	del grupo de 533 genot.
11124	0. pibentima	del grupo de 533 genot.
121240	0. pibentima	del grupo de 533 genot.
1310000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
1400000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
1500000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
1600000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
1700000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
1800000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
1900000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
2000000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
2100000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
2200000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
2300000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
2400000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
2500000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
2600000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
2700000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
2800000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
2900000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
3000000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
3100000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
3200000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
3300000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
3400000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
3500000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
3600000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
3700000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
3800000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
3900000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
4000000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
4100000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
4200000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
4300000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
4400000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
4500000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
4600000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
4700000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
4800000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
4900000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
5000000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
5100000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
5200000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
5300000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
5400000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
5500000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
5600000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
5700000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
5800000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
5900000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
6000000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
6100000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
6200000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
6300000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
6400000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
6500000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
6600000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
6700000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
6800000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
6900000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
7000000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
7100000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
7200000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
7300000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
7400000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
7500000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
7600000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
7700000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
7800000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
7900000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
8000000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
8100000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
8200000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
8300000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
8400000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
8500000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
8600000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
8700000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
8800000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
8900000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
9000000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
9100000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
9200000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
9300000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
9400000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
9500000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
9600000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
9700000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
9800000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
9900000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.
10000000	1805144-28-2-2-1-1-100	Hijos de genot. Con am. Fe.

B

Selecciones y Líneas avanzadas BG90-270, rullipogon
49 Materiales - CIAT - 2003

Cont.	Parcela	Origen	Pedigree	GA3	FE	FE	FE	Long	%	Pres
Cont.	Parcela	Origen	Pedigree	GA3	FE	FE	FE	Long	%	Pres
1	85	216	CT13088-13-M-2-4-M-3-11	108	48	34	20	17	20	
2			CT13088-13-M-2-4-M-3-12	42	38	23	19	19	24	
3			CT13088-13-M-2-4-M-3-13	48	35	25	18	18	24	
4			CT13088-13-M-2-4-M-3-14	51	24	22	18	18	24	
5			CT13088-13-M-2-4-M-3-15	81	39	24	24	24	28	
6			CT13088-13-M-2-4-M-3-16	52	38	23	17	17	20	
7			CT13088-13-M-2-4-M-3-17	87	20	22	21	21	20	
8			CT13088-13-M-2-4-M-3-18	103	39	24	15	15	21	
9			CT13088-13-M-2-4-M-3-19	100	37	34	20	20	25	
10			CT13088-13-M-2-4-M-3-20	87	30	22	22	22	25	
11	81	80	CT13088-29-M-29-2-4-M-3-11	108	85	39	21	24	44	
12			CT13088-29-M-29-2-4-M-3-12	88	35	22	22	22	27	
13			CT13088-29-M-29-2-4-M-3-13	85	39	22	18	18	27	
14			CT13088-29-M-29-2-4-M-3-14	85	39	22	18	18	27	
15			CT13088-29-M-29-2-4-M-3-15	87	35	21	30	30	29	
16			CT13088-29-M-29-2-4-M-3-16	80	35	23	30	30	29	
17			CT13088-29-M-29-2-4-M-3-17	83	18	21	47	47	25	
18			CT13088-29-M-29-2-4-M-3-18	86	15	21	27	27	25	
19			CT13088-29-M-29-2-4-M-3-19	84	30	21	30	30	38	
20			CT13088-29-M-29-2-4-M-3-20	88	31	21	34	34	38	
21	80	102	CT13088-29-M-29-2-4-M-3-21	113	42	23	30	30	47	
22			CT13088-29-M-29-2-4-M-3-22	86	17	21	51	51	41	
23			CT13088-29-M-29-2-4-M-3-23	85	22	19	51	51	57	
24			CT13088-29-M-29-2-4-M-3-24	82	14	21	50	50	57	
25			CT13088-29-M-29-2-4-M-3-25	88	19	20	52	52	57	
26			CT13088-29-M-29-2-4-M-3-26	90	21	19	58	58	61	
27			CT13088-29-M-29-2-4-M-3-27	82	18	19	48	48	58	
28			CT13088-29-M-29-2-4-M-3-28	80	22	20	51	51	57	
29			CT13088-29-M-29-2-4-M-3-29	88	31	22	55	55	58	
30			CT13088-29-M-29-2-4-M-3-30	87	40	21	58	58	59	
31			CT13088-29-M-29-2-4-M-3-31	85	22	19	43	43	57	
32			CT13088-29-M-29-2-4-M-3-32	88	24	20	57	57	59	
33			CT13088-29-M-29-2-4-M-3-33	89	23	21	57	57	59	
34			CT13088-29-M-29-2-4-M-3-34	88	47	24	57	57	59	
35			CT13088-29-M-29-2-4-M-3-35	87	37	20	52	52	59	
36			CT13088-29-M-29-2-4-M-3-36	88	24	20	57	57	59	
37			CT13088-29-M-29-2-4-M-3-37	83	20	21	51	51	59	
38			CT13088-29-M-29-2-4-M-3-38	87	3	19	58	58	59	

C

Primer	Secuencia	%GC	Tm (Ta)
Ibpps 1.2 Fw	CCATCAAGGCACCTACCGGA	55	57
Ibpps 2 Fw	GTATGTTTGTAAACAATAGGG	35	46
Ibpps 1.2 Rv	ACTTGAGCTGAGGTGAATA	40	52
5'UTR 1	TTGATTGTTGTTGGGGGTTT	38	53
5'UTR 2	TTGATTGTTGTTGTTTATTG	29	51
5'UTR 3	CGGAGAAAGCGAGAGCGTTA	55	57
5'UTR 4	AAATTGACACGTTTGCGCAT	38	54
5'UTR 5	GTATTATGTCCTAAGCGCCAA	43	56

Figura 30. ejemplos de listados en formato electrónico: a) listado del material vegetal, b) datos fenotípicos importados de libros de campo, y c) listado de cebadores.

- Resultados parciales. Los resultados parciales encontrados en formato digital se presentaron únicamente a manera de párrafos cortos (figura 31).

PCR-ITS (Concentración final de los primers [0.1 micromolar])
mejoro mucho!!!!!!!!!!!!

Figura 31. resultado parcial presentado en formato digital.

- Como se hace (How-to). Instrucciones sobre como hacer ciertos procedimientos de laboratorio también fueron importados de diferentes medios electrónicos y pegados en los cuadernos (figura 32).

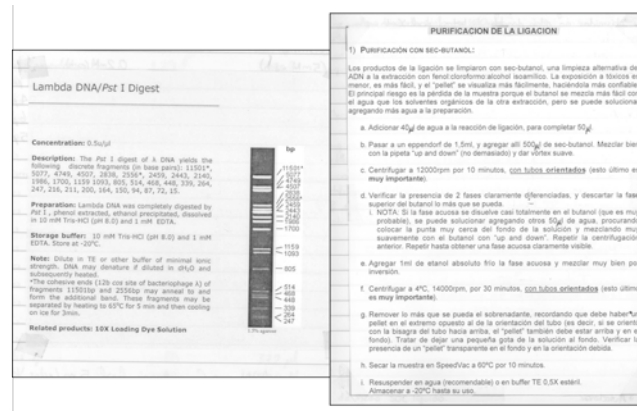


Figura 32. Guías How-to, importadas de medios electrónicos a los cuadernos de laboratorio.

- Figuras. Representación dibujada o gráfica de un cuerpo, objeto, molécula, etc. encontrada en los cuadernos de laboratorio (figura 33).

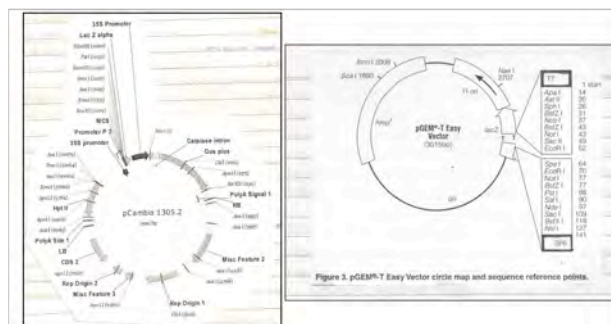


Figura 33. Representación grafica de vectores, son ejemplos de figuras encontradas en los cuadernos analizados.

- Secuencias parciales. Aquí se define secuencia parcial como una sucesión de letras representando la estructura primaria de una molécula real o hipotética de algún tipo de ácido nucleico, con la capacidad de transportar información. En algunos cuadernos de laboratorio se encontraron algunas secuencias parciales de ácidos nucleicos importada de bases de datos disponible en la Web (figura 34).



CCATCAAGGCACCTACCGGATCTGTCTCAAAGATGAAAATCTACAGGCTCTTTATAGGACTCTAGCCACTGTCATCA
 CCTTTAACTTTAAGACAAAAGTGACGGTGACAGATATTGATTCTTGCACTAAGACACAACAATTGTTCTCTGACATT
 ATTTCTAACTTCTAAGATAAAAGTAACCGTTCAAATATTGATTCTTTGGTTGAAACACGTGAACCTCACTGCTCACTT
 CTAACCTCTAAAACAAAAGTAACCGTTCAATATTAAATTATGTTATCAGGAAATACGCAAGCTTTTCATATACTGGTT
 TATTGCTCTAAGGCGGCAATCATAATTCTGTTTTCATCTTATGACGATTATAATTGTTTTAACAGAAAATTAATAAA
 GAAATTGTACACTTTTAATCAAATCATCGCCACATTATGTTATTTAGACTCAATTACACTATTATTGATTAAAT
 TTATTAATAATGTATAATATGAAACTAAATTGACACGTTTGGGCATTAAATAGGACTAGCTGAATTAACCAACA
 CGAATAATGTCATGAAGGCGTGGGTTGAAGAGACGCAATCGCCAACAGTATCCAAAAGCAAAGACCGTTTGCCACGT
 TTCAAACGCAGGAGGCGCCACGTATCCACTGCATGACAGCTAGACACGAAACCGAGAGAAAGCGAGAGCGTTAGATAT
 AGGGAGAGGCGACTGTTGGGCGATAATCTTGCAGATTATTTATGGGTTTCTTTTCTCAATATTAAGATCCTT
 TATTCTGCAGCACTGTTGGGAAGTGCTGTTGTTGATTAAAGCTTTTGTATCTCTACTGATTTTGTGTTAAAAGGG
 CATATGATGCTATAATTGAGTAAGCACAGCAGGTTACTTAGTTTGATTGTGGTTCAATTCATTGTCAGAAAGGAAATA
 AAGAGGGGAGCTTCTGTTATAATTCAATTCATGAACCTTTGCTATTTGTGGCAGTTAAAGCAAAGTCAGTTTG
 ATTTGGTTTGGGGGTTAATAATCTTTTGCAAAAAGCCAAAGTGGGCATTGTTGAAAAGTGGATAGGCTTTAGT
 GGGCATATTCATTTTACTATAAACACACAGGATAATCAAGGGGACAGAAATACAAAATCTTGGGAAGTAGTTATAT
 AGAAACTATTACCTCAGCTCAAGT

Figura 34. tipo de secuencia parcial encontrada en los cuadernos de laboratorio.

Los seis tipos de información digital identificados en este estudio se resumen en la tabla 2. Estos datos, al igual que los tipos de información manuscrita, fueron organizados de mayor a menor número de ocurrencias en las 408 páginas analizadas. El total de datos manuscritos encontrados fue de 110.

Tabla 2. Tipos de información digital identificada en los cuadernos de laboratorio analizados.

Información digital		información importada de herramientas externas	
	fotografías	Ácidos nucleicos cargados en geles de agarosa	51
		Ácidos nucleicos cargados en geles de poliacrilamida	11
		Total	62
	listados	material vegetal Vs. características agronómicas	11
		material vegetal Vs cebadores	5
		material vegetal	3
		cebadores	7
		Total	26
	resultados parciales	párrafos describiendo los resultados obtenidos	11
	Como se hace (how-to)	ensayos de campo	1
		Lambda DNA/Pst I Digest	1
		eliminación de bandas inespecíficas	1
		purificación de ligación	1
		extracción de ácidos nucleicos	1
		posición de las muestras en placas de PCR	3
		Total	8
	figuras	vectores de expresión	2
	secuencias parciales	DNA, RNA.	1

110

3.4.3. Información Digital Conteniendo Anotaciones Manuscritas. Se encontraron tablas elaboradas en formato excel o word usadas como plantilla de trabajo. En dichas plantillas se agregó de manera manuscrita, información requerida en algunos tipos de procesos experimentales conservados en los cuadernos de laboratorio. De igual manera se encontraron anotaciones manuscritas sobre algunas fotografías de geles de agarosa con el propósito de identificar cada una de las muestras de ADN allí contenidas. Finalmente se encontró un caso de anotación manuscrita sobre una secuencia parcial de ADN. La figura 35 ilustra los dos tipos mas comunes de información digital conteniendo anotaciones manuscritas.

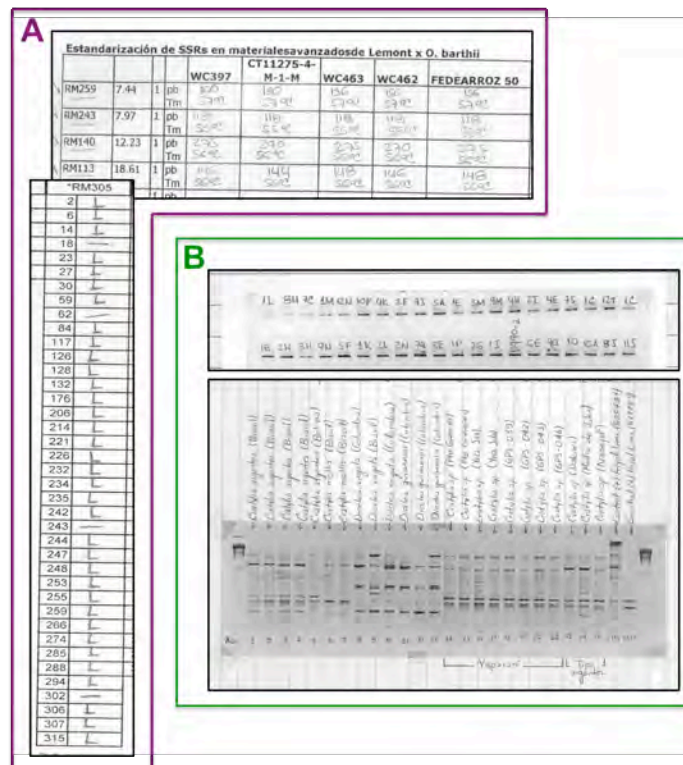


Figure 35. A) plantillas de trabajo diseñadas para diferentes actividades experimentales. B) fotografías de geles de agarosa conteniendo anotación manual.

La tabla 3 resume los tres tipos de información digital conteniendo información manuscrita encontrados en este estudio.

Tabla 3. Tipos de información digital conteniendo anotación manuscrita identificada en los cuadernos de laboratorio analizados.

información manuscrita + digital		
anotaciones manuscritas hechas sobre:	fotografías	51
	secuencias parciales	1
	tablas excel	7
		59

Luego de analizar en detalle cada uno de los tres grupos principales de datos: 1) Información manuscrita, 2) información digital, e 3) información digital con anotaciones manuscritas; se pudo elucidar lo siguiente:

- El grupo de datos conocido como “Información manuscrita creada y editada directamente en los cuadernos de laboratorio”, fue el grupo de datos mas popular, con una frecuencia de registro del 93.02% (figura 37).
- De manera conjunta, los grupos de datos “información digital” e “información digital con anotaciones manuscritas” representan solo el 6.98% del total de datos conservados en las paginas de cuadernos de laboratorio analizadas (figura 38).
- El subtipo de dato identificado como “verificación de cada paso o actividad realizada” cuenta con el mayor numero de registros (1288 \approx 53.20%) (figura 37). La forma mas común usada para verificar cada paso o actividad realizada fue mediante el uso de un visto bueno (\checkmark). Este símbolo también resultó ser un indicador del numero de veces que fue realizada una determinada actividad experimental bajo las mismas condiciones (figura 36). Esta medida adoptada por el investigador constituyó una manera de evitar la reescritura de procedimientos de rutina.

	12x	10 2x
ADN	5ul	
B. 10x lab y con	2ul	1111 20ul
Mgel2	12ul	1111 12ul
distes	0,3ul	1111 3ul
PRIMEE R	0,1ul	1111 1ul
PRIMEE F	0,1ul	1111 1ul
Taq lab y con	0,1ul	1ul
H ₂ O	11,2ul	1111 112ul
V.F = 20ul		

Figura 36. se ilustra el numero de veces (4) que fue realizada una PCR bajo las mismas condiciones experimentales.

- Los subtipos de datos identificados como “ruta de archivos electrónicos”, “fuente de la información” y “secuencias parciales” fueron los menos comunes, cada uno con 1 registro (0.04%) (figuras 37 y 38).

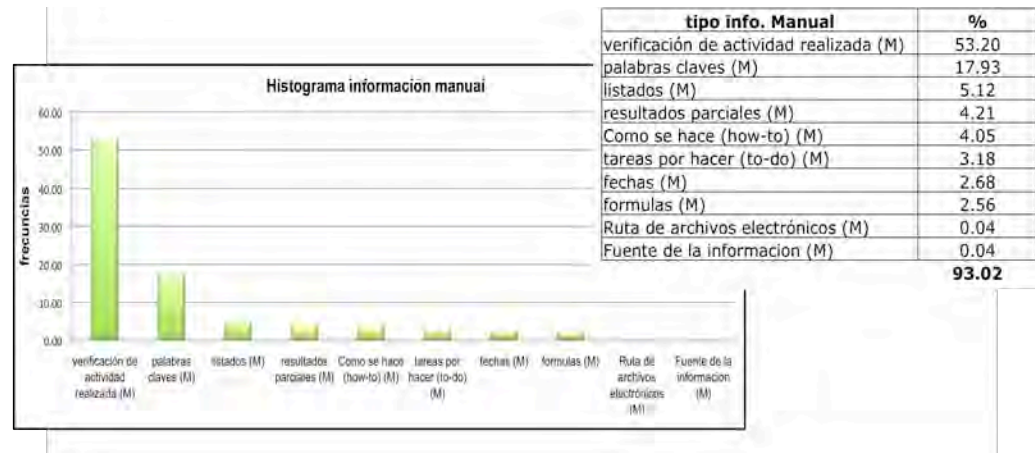


Figura 37. histograma que representa la frecuencia de registro de cada uno de los 10 subtipos de información manual.

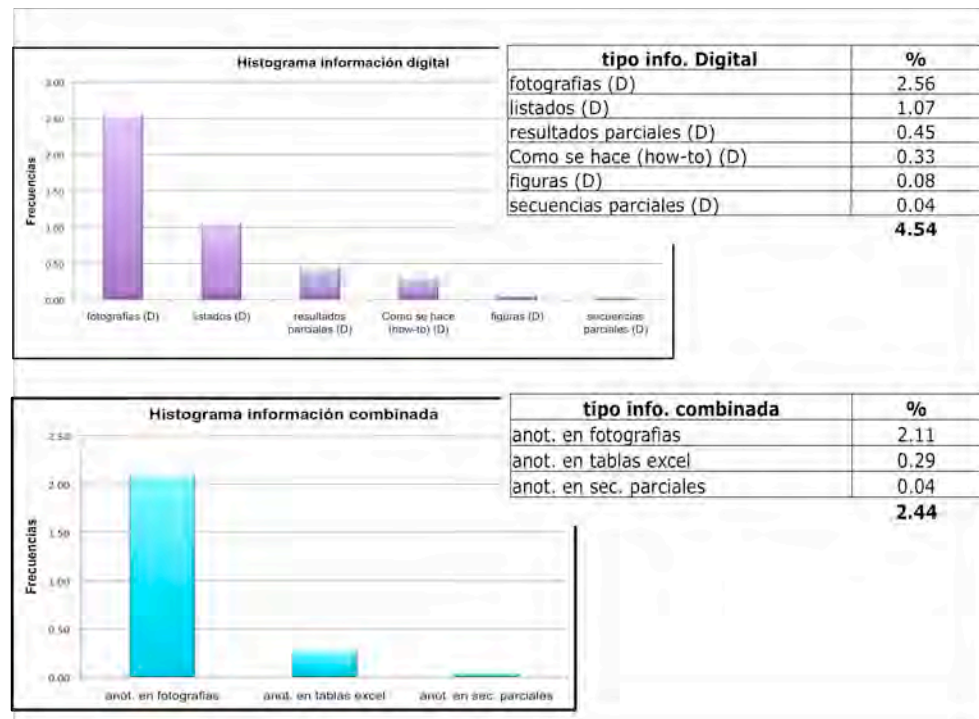


Figura 38. histograma que representa la frecuencia de registro de los 6 subtipos de información digital y la frecuencia de registro de los 3 subtipos de información digital con anotaciones manuscritas.

Los resultados anteriores demuestran que la capacidad de los cuadernos de laboratorio se limita al número de hojas y su versatilidad para interactuar con medios electrónicos. Por ello, Los investigadores no pueden consignar toda la información que producen en estos dispositivos; teniendo que recurrir a medios alternos como computadoras personales, dispositivos USB, CDs, DVDs, medios físicos como carpetas, etc.

3.5. METADATOS

La falta de un índice o tabla de contenido hace que el proceso de recuperación de información se soporte en la capacidad de memoria del investigador. La poca procesabilidad de datos manuscritos, no estructurados y sin anotación dificulta el uso de un recurso informático, dicho recurso, es una alternativa diferente a la dependencia que se tiene de las capacidades cognitivas del dueño del cuaderno de laboratorio. Con el proposito de facilitar la organización de la información se propone aquí el uso de metadatos. Estos describen la información contenida en cualquier tipo de recurso, haciendo posible algún procesamiento sistemático de la misma. Con el propósito de ilustrar esta alternativa se presenta un ejemplo.

Ejemplo

¿Que hay dentro de estos cuadernos, ver figura 39, que tipo información contienen?



Figura 39. Cuadernos de laboratorio no auto-descriptivos.

Ninguno de los cuadernos “se presenta a si mismo”; no contienen información auto-descriptiva. Siendo éstos, repositorios de actividades experimentales, una marca obvia sería aquella que indicara cuando y donde una actividad experimental se inicia y continúa. La falta de marcas hace necesaria la inspección manual de los tres cuadernos. Una pregunta un tanto mas compleja como: “se tienen consignados en estos cuadernos, experimentos relacionados con tolerancia a salivazo en *Brachiaria*?” implicaría un trabajo manual de minería de datos para poder determinar donde se consignó información relativa a la tolerancia a salivazo en *Brachiaria*.

Soluciones sencillas, como por ejemplo el uso de etiquetas, pueden proporcionar elementos de auto-descripción, ver figura 40; el facilitar índices también ayuda. En ambos casos se requiere una estructuración de la información, además de consistencia en el uso de los mismos descriptores.

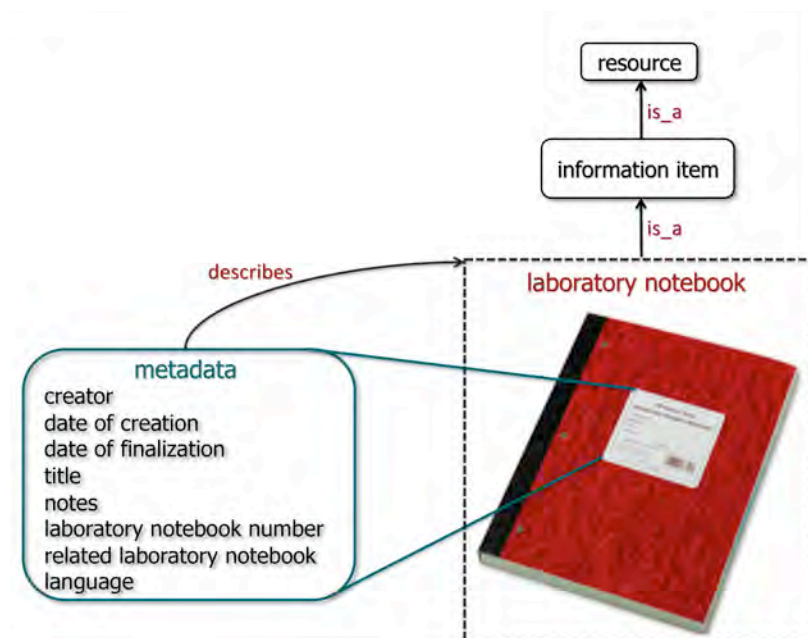


Figura 40. Cuaderno auto-descriptivo.

3.5.1 Usando los Metadatos. Los metadatos se definen como “*datos sobre los datos*” o “*lo que está más allá de los datos*”. Esta definición, aunque comúnmente aceptada, aporta muy poco significado (Murphy, 1998, Paganelli et al., 2006, Päivärinta et al., 2002). En el ámbito de la gestión de la información, el término metadato se emplea para referirse a registros descriptivos de recursos digitales disponibles en la red (Heery, 1996). Los metadatos son utilizados para referenciar de manera estandarizada cualquier información acerca de recursos digitales y no digitales (Steinacker et al., 2001). Algunos ejemplos del uso de metadatos se presentan en la tabla 4.

Tabla 4. Metadatos asignados a diferentes objetos.

“Datos sobre los datos”	Objeto
Documento de identidad Nombre, numero de identificación, huella digital, tipo de sangre, etc.	Una persona
Etiqueta de un envase Foto del contenido, fecha de vencimiento, contenido nutricional, fabricante, etc.	Contenido del envase
Ficha bibliográfica Titulo, autor, ISBN, DOI, tema, etc.	Libro

Los metadatos pueden servir para resumir el contenido de la fuente de información, permitir su búsqueda y recuperación, especificar la autoria o responsabilidad, especificar las condiciones de uso (en particular los derechos de autor), aportar información sobre la vida útil del recurso o especificar las relaciones existentes entre otros, etc. (Ercegovac, 1999).

3.5.2 Aplicación de Metadatos Normalizados. Existen varios modelos de metadatos, uno de ellos es Dublin Core, un simple pero eficaz conjunto de elementos para describir una amplia gama de recursos. Existe también AgMes, un set de metadatos usados para la descripción de recursos agrícolas. Igualmente se tiene e-GMS (e-Government Metadata Standard) y AGLS, metadatos normalizados aplicados a recursos gubernamentales. A continuación se describen los cuatro modelos de metadatos antes mencionados.

- Dublin Core. El Dublin Core (DC), ha sido el modelo mas difundido y utilizado (Paganelli et al., 2006, Kakali et al., 2007, Murphy, 1998, Kimura et al., 2009). Originalmente se aplicaba sobre el lenguaje de marcado HTML (HyperText Markup Language -*Lenguaje de Marcado de Hipertexto*); HTML, se usa para describir la estructura y el contenido en forma de texto, así como para complementar el texto con objetos tales como imágenes (Beatty et al., 2008). Actualmente el DC, puede utilizarse en diferentes entornos tecnológicos. La semántica del DC se ha establecido por un grupo internacional e interdisciplinario de profesionales de la biblioteconomía, la Informática, la codificación textual, la comunidad museística, y otros campos teórico-prácticos relacionados (Weibel, 2000).

- AGLS. El AGLS¹⁴ es un conjunto de 19 elementos descriptivos que los departamentos gubernamentales y agencias pueden utilizar para mejorar la visibilidad y la accesibilidad de sus servicios e información a través de Internet. El Archivo Nacional de Australia es la agencia que mantiene la normativa de metadatos AGLS.

¹⁴ AGLS Metadata Standard, available from <http://www.agls.gov.au/>

- AgMes metadata element set. AgMes¹⁵ es el set de metadatos estándar de la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO). Su utilidad se deriva de facilitar metadatos que son importantes para la descripción de los recursos de información en el ámbito de la Agricultura. AgMes es un espacio de nombres, definición de elementos y refinamientos que se consideran como extensiones de las normativas DC y AGLS.

- e-GMS (e-Government Metadata Standard). El e-GMS¹⁶, establece los elementos, refinamientos y esquemas de codificación que pueden ser utilizados por organismos del sector publico del Reino Unido. El propósito de este conjunto de metadatos es facilitar la recuperación y gestión de información.

3.5.3 Modelo para la Representación de Conceptos. Para la aplicación de metadatos se han desarrollado distintos modelos capaces de proporcionar los elementos necesarios para establecer relaciones entre los metadatos (términos, conceptos). Uno de los modelos mas usados, SKOS, se define a continuación (Cantara, 2006, Gray et al., 2008, Dasgupta and Rath, 2009).

¹⁵ AgMes metadata element set. Available from, <http://aims.fao.org/en/agmes-metadataset>.

¹⁶ (2006) e-Government Metadata Standard Version 3.1. e-Government Unit, Stockley House, 130 Wilton Road, London SW1V 1LQ, govtalk@cabinet-office.x.gsi.gov.uk.

- SKOS (Simple Knowledge Organization System). SKOS¹⁷, es un modelo usado para la representación de sistemas de organización del conocimiento (Knowledge Organization Systems, SKOS) semi-formales, tales como tesauros, taxonomías, esquemas de clasificación entre otros. SKOS está basado en un lenguaje RDF (Resource Description Framework) por lo que dichas representaciones pueden ser legibles por máquinas e intercambiarse entre aplicaciones de software, así como publicarse en la World Wide Web. SKOS ha sido diseñado para proporcionar un modo de migrar a la Web Semántica sistemas de organización del conocimiento ya existentes con un bajo costo. SKOS también proporciona un lenguaje conceptual de modelado muy sencillo e intuitivo para desarrollar y compartir nuevos sistemas de organización. Puede utilizarse solo, o en combinación con lenguajes más formales, como el Lenguaje de Ontologías Web (OWL). SKOS también puede contemplarse como una tecnología de transición que proporciona un nexo entre el formalismo lógico riguroso de los lenguajes de ontologías como OWL y el mundo caótico, informal y débilmente estructurado de las herramientas colaborativas basadas en Web, ejemplificadas por las aplicaciones de etiquetado social. El objetivo de SKOS no es sustituir vocabularios originales en su contexto inicial de uso; es en cambio su objetivo el que puedan implementarse en un espacio compartido, basado en un modelo simplificado, que haga posible su reutilización y una mejor interoperabilidad.

3.5.4 Metadatos para la Identificación de Cuadernos de Laboratorio. La tabla 5 presenta un conjunto de metadatos derivados del análisis de las normativas DC, AGLS, AgMes y e-GMS; en este análisis se contempló como facilitar la organización y clasificación de los cuadernos de laboratorio.

¹⁷ SKOS. 2003; Available from: <http://www.w3.org/2004/02/skos/>.

Tabla 5. Conjunto de metadatos que soportan la auto-descripción del cuaderno de laboratorio.

Fuente	Nombre del metadato	Definición	Propósito
DC AgMes	ET: creator. ST: creador	ED: Person, organization or a service primarily responsible for creating or making contributions to the intellectual content of the resource. SD: Persona, organización o servicio responsable de crear o de hacer contribuciones al contenido intelectual del recurso.	* Revela quien o quienes son los responsables. * Búsqueda por atributo.
DC	ET: date of creation. ST: fecha de creación.	ED: Date of creation of a resource. SD: Fecha de creación de un recurso.	* Control y manejo. * Búsqueda por atributo.
DC M4L	ET: date of finalization. ST: fecha de finalización.	ED: Date of finalization of a resource. SD: Fecha de finalización de un recurso.	* Control y manejo. * Búsqueda por atributo.
DC	ET: title. ST: título.	ED: A name given to the resource. SD: Nombre dado a un recurso.	* Búsqueda por atributo.
AgMes	ET: notes. ST: notas.	ED: Brief statement, annotation, comment, or/and elucidation concerning any aspect of the resource. SD: Breve declaración, anotación, comentario, y/o aclaración sobre cualquier aspecto del recurso.	* Evaluación de utilidad y relevancia. * Búsqueda por atributo.
M4L	ET: laboratory notebook number. ST: número del cuaderno de laboratorio.	ED: Laboratory notebook number is a number allotted to a laboratory notebook for purposes of identification and retrieval. SD: Número de cuaderno de laboratorio es un número asignado a un cuaderno de laboratorio para fines de identificación y recuperación.	* Control y manejo. * Búsqueda por atributo.
DC	ET: language. ST: lenguaje.	ED: A language of the resource. SD: Lenguaje del recurso.	* Por información. * Búsqueda por atributo.

ET= English Term, ST=Spanish Term, ED=English Definition, SD=Spanish Definition.

Como se observa en la tabla 5, siete metadatos fueron seleccionados para describir un cuaderno de laboratorio. Cinco de ellos provienen de la normativa DC¹⁸, dos fueron seleccionados del set de metadatos AgMes¹⁹. Los elementos extraídos de DC y AgMes, sirven para describir cualquier tipo de recurso (artículo, reporte, actividad experimental, cuaderno de laboratorio, etc.).

Con el propósito de identificar el conjunto de metadatos propuestos en esta tesis, se utilizará el nombre M4L –Metadata for Laboratory Notebook. M4L contiene, metadatos completamente nuevos y propios del problema de estudio, además de aquellos metadatos producto del refinamiento de algunas propuestas disponibles. Las anotaciones hechas por los autores de los cuadernos estudiados se depuraron, encontrando para ellas niveles mas altos de abstracción –permitiendo así su mejor organización. Este conjunto de metadatos, tiene como propósito el describir un cuaderno de laboratorio, independientemente de los datos que este pueda almacenar.

3.5.5 Metadatos para Descripción de una Actividad Experimental. Para la selección de elementos descriptivos de una actividad experimental cualquiera, sin datos, se revisaron las normativas DC, AgMes y AGLS. Además, se complementó esta revisión con glosarios, tesauros y ontologías descritos a continuación.

¹⁸ *Dublin Core Metadata Element Set, Version 1.1.* 2008; Available from: <http://dublincore.org/documents/dces/>.

¹⁹ *AgMes metadata element set.* Available from: <http://aims.fao.org/en/agmes-metadataset>.

- NAL Agricultural Thesaurus (NALT). El Tesauro Agrícola fue preparado en primera instancia por la Biblioteca Nacional de Agricultura (NAL, por sus siglas en inglés) para resolver las necesidades del Departamento de Agricultura de EE.UU (USDA, por sus siglas en inglés) y del Servicio de Investigación Agrícola (ARS, por sus siglas en inglés). La primera edición del Tesauro Agrícola fue publicada el 1 de enero de 2002 y se actualiza en su formato digital de manera anual (National Agricultural Library (NAL) and Agricultural Research Service, 2010b).

- Glossary of Agricultural Terms. El Glosario Agrícola, es una colección de definiciones de términos agrícolas, desarrollado paralelamente con la creación del Tesauro promovido por de la Biblioteca Nacional de Agricultura. La edición del año 2009 contiene más de 2,649 definiciones para los temas de agricultura y sus asuntos complementarios (National Agricultural Library (NAL) and Agricultural Research Service, 2010a).

- AGROVOC. AGROVOC²⁰ es un tesauro estructurado multilingüe que abarca todos los ámbitos de la agricultura, la silvicultura (ciencias forestales), la pesca, alimentos y otros ámbitos afines (por ejemplo, el medio ambiente). Ha sido creado y mantenido por la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO), ha estado disponible desde principios de los años ochenta. AGROVOC nació como un diccionario de sinónimos, ha

²⁰ AGROVOC: from a thesaurus to an ontology. Available from:
<http://aims.fao.org/website/AGROVOC%20Concept%20Server/sub>.

evolucionando hacia una ontología; se compone de palabras en diversos idiomas, y relaciones tales como "más amplio", "término específico", "afines", etc. Su función principal es normalizar el proceso de indexación a fin de que las búsquedas sean más simples y eficaces, y para proporcionar al usuario los recursos más relevantes.

- BioPortal Metadata. BioPortal Metadata, contiene diferentes tipos de metadatos para describir las ontologías que se guardan en el repositorio BioPortal . Este conjunto de metadatos se deriva del Ontology Metadata Vocabulary (OMV) (Hartmann et al., 2005)

- NCI Thesaurus (NCIt). La normativa NCIt es ampliamente reconocida en el área biomédica; proporciona terminología de referencia para varios centros de investigación que hacen parte del National Cancer Institute. El NCIt, es además utilizado por una amplia variedad de socios públicos y privados tanto a nivel nacional e internacional tales como: Clinical Data Interchange Standards Consortium Terminology (CDISC), la U.S. Food and Drug Administration (FDA), la Federal Medication Terminologies (FMT), y el National Council for Prescription Drug Programs (NCPDP). El vocabulario cubre áreas relacionadas con atención clínica, investigación, información pública y actividades administrativas. Las características del NCIt son:

- Estable, presenta códigos exclusivos de conceptos biomédicos;

- Incluye, términos, sinónimos, definiciones, códigos de investigación, códigos de fuente externa;
- Permite hacer enlaces a Meta-tesauros del NCI y otras fuentes de información;
- Se actualiza con frecuencia por un equipo de expertos en la materia.

- SNOMED-CT. SNOMED-CT²¹ (Systematized Nomenclature of Human and Veterinary Medicine- Clinical Terms) es una estructura de codificación mantenida por el Colegio Americano de Patólogos (CAP) y está ampliamente aceptada para describir los resultados de pruebas clínicas. Tiene una estructura de codificación multi-axial con once campos lo que le confiere especificidad, dándole un considerable valor para fines clínicos.

- The Ontology for Biomedical Investigations (OBI). El proyecto OBI²² es producto de un esfuerzo colaborativo internacional que busca la construcción de una serie de ontologías para la descripción de investigaciones clínicas y biomédicas. La ontología incluye un set de términos universales que pueden aplicarse en varias áreas del dominio biológico y tecnológico; también cuenta con términos específicos a algunos dominios. A partir de esta ontología es posible representar el diseño de una investigación, los protocolos, los instrumentos, el material usado, los datos generados y el tipo de análisis que se requiera.

²¹ SNOMED CT. Available from: <http://www.ihtsdo.org/snomed-ct/>.

²² The Ontology for Biomedical Investigations. Available from: http://obi-ontology.org/page/Main_Page.

3.5.6 Presentación de los Metadatos que Autodescriben una Actividad Experimental. El producto del análisis de los anteriores vocabularios controlados y metadatos se presenta en la tabla 6. Aquí se definen los términos para describir una actividad experimental.

Tabla 6. Conjunto de metadatos que soportan la auto-descripción de una actividad experimental.

Fuente	Nombre del metadato	Definición	Propósito
OBI AGROVOC	ET: project. ST: proyecto.	ED: Set of laboratory procedures coordinated and interrelated that seek to comply with some specific goal. This usually has to be reached in a predefined time period and maintaining a budget. SD: Conjunto de procedimientos de laboratorio coordinados e interrelacionados que buscan cumplir con un cierto objetivo específico. Este generalmente debe ser alcanzado en un periodo de tiempo previamente definido y respetando un presupuesto.	* Búsqueda por atributo.
M4L NCIt	ET: laboratory procedure title. ST: título del procedimiento de laboratorio.	ED: A name given to a test or experimental activity. SD: Nombre dado a una prueba o actividad experimental.	* Por información. * Búsqueda por atributo.
DC	ET: date. ST: fecha.	ED: Element used to designate the date associated with an event in the life cycle of the resource. Typically, Date will be associated with the creation or availability of the resource. SD: Elemento utilizado para designar la fecha asociada a un evento en el ciclo de vida del recurso. Normalmente, la fecha se asociará con la creación o la disponibilidad del recurso.	* Control y manejo. * Búsqueda por atributo.
M4L	ET: recorded by ST: registrado por.	ED: Person responsible for record the experimental activity. SD: Persona responsable del registro de la actividad experimental.	* Control y manejo. * Búsqueda por atributo.

ET= English Term, ST=Spanish Term, ED=English Definition, SD=Spanish Definition.

Fuente	Nombre del metadato	Definición	Propósito
M4L	ET: page number. ST: número de pagina.	ED: Page of laboratory notebook where one person currently is recording the experimental activity. SD: Página del cuaderno de laboratorio donde una persona actualmente se encuentra grabando la actividad experimental.	* Control y manejo. * Búsqueda por atributo.
M4L	ET: purpose. ST: propósito.	ED: Element used to include a brief description of the usefulness of the test to be performed. SD: Elemento usado para incluir una breve descripción de la utilidad de la prueba que se va a realizar.	* Evaluación de la utilidad y relevancia.
M4L	ET: security measurements. ST: medidas de seguridad.	ED: Element used to describe the security measures required, depending on the characteristics of the test and reagents. SD: Elemento utilizado para describir las medidas de seguridad requeridas en función de las características de la prueba y los reactivos.	* Por información. * Búsqueda por atributo.
OBI	ET: protocol. ST: protocolo.	ED: A protocol is a plan specification which has sufficient level of detail and quantitative information to communicate it between domain experts, so that different domain experts will reliably be able to independently reproduce the process. SD: Un protocolo es una especificación planeada que tiene suficiente nivel de detalle y la información cuantitativa para comunicarla entre los expertos de dominio, de modo que diferentes expertos de dominio de forma fiable será capaz de reproducir el proceso de forma independiente.	* Evaluación de la utilidad y relevancia.
NCIt	ET: outcome ST: salida/resultado	ED: The result of an action. In science, the <i>outcome</i> of an experimental activity. ST: El resultado de una acción. En ciencia, el resultado de una actividad experimental.	• Por información.
BioPortal NCIt SNOMED	ET: comments. ST: comentarios.	ED: Element used to make a comment of experimental activity results. ST: Elemento utilizado para hacer un comentario de los resultados de la actividad experimental.	* Por información.

ET= English Term, ST=Spanish Term, ED=English Definition, SD=Spanish Definition.

En la tabla 7 se presenta un total de diez metadatos. Uno de ellos “Date”, fue tomado de la normativa DC. El término “Project” es usado por las ontologías AGROVOC y OBI, pero la definición es propia de este estudio. El elemento “Laboratory Procedure Title” es un refinamiento derivado del tesauro NCI. Tres elementos (Protocol, Outcome y Comment), fueron tomados de OBI (Protocol), del tesauro NCI (Outcome) y el término “comment” esta contenido, bajo la misma connotación, en BioPortal Metadata, SNOMED-CT y el tesauro NCI. Por último, cuatro nuevos elementos entraron a formar parte del conjunto de metadatos propuestos en esta tesis (M4L), los cuales son específicos para soportar la auto-descripción de una actividad experimental.

Este conjunto de metadatos, es el resultado de analizar el tipo de datos que el investigador usa para describir cualquier procedimiento experimental y de consultar información existente sobre el como llevar un cuaderno de laboratorio (Kanare, 1985). La figura 41, se presenta un mapa conceptual que contiene los dos conjuntos de metadatos que describen un cuaderno de laboratorio y un procedimiento experimental.

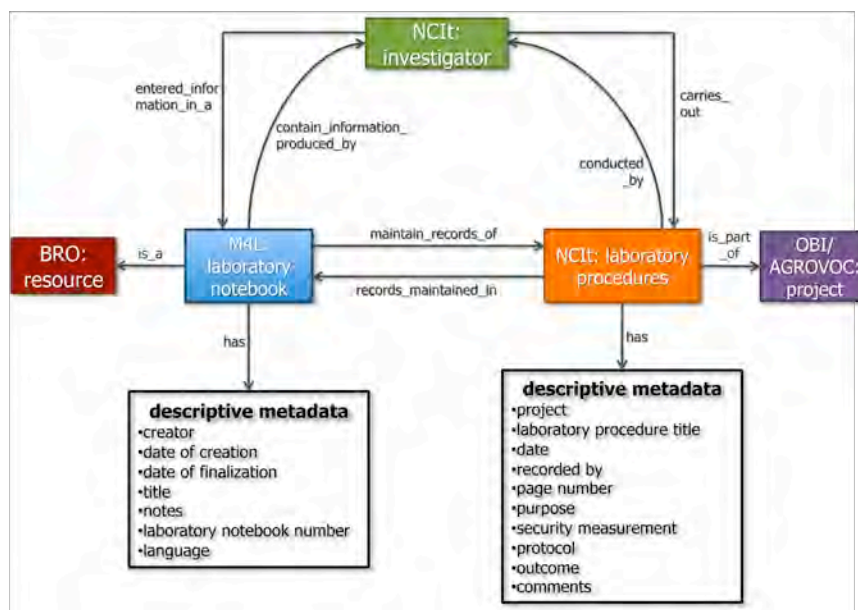


Figura 41. Presentación de los metadatos que autodescriben un cuaderno de laboratorio y un procedimiento experimental.

3.6 COMENTARIOS FINALES

En este capítulo se analizaron 12 cuadernos de laboratorio; sobre la base del análisis hecho, se derivaron metadatos coherentes a partir de conjuntos de metadatos existentes. Se consideró el cuaderno como objeto físico y el experimento como una actividad; en ambos casos, independientemente de los datos almacenados en el cuaderno y generados por la actividad experimental.

La falta de estructura, escritura en ocasiones ilegible, falta de detalle en la manera de presentar la información, la inexistencia de índices y el uso de datos o anotaciones no formales para la descripción tanto del cuaderno de laboratorio como de una actividad experimental fue evidente en los cuadernos analizados. La principal consecuencia, dificultad en la búsqueda de información pertinente y por ende, deficiencia en el proceso de recuperación de la información.

Una de las ventajas de llevar un cuaderno de laboratorio, es el uso de un lenguaje natural o libre, sin embargo, esto en ocasiones conduce a la sinonimia o ambigüedad y con ello se incrementa significativamente la cantidad de información recuperada, pero reduce la precisión o relevancia de la información que se recupera. Por esta razón, en este estudio se propuso, el uso de metadatos normalizados extraídos de sets de metadatos estándar, glosarios, tesauros y ontologías para ofrecer una mayor consistencia a las palabras utilizadas, la correcta indización tanto del cuaderno de laboratorio como de cualquier actividad experimental, y eliminar los problemas de ambigüedad y sinonimia que tanto afectan la recuperación de la información.

En síntesis, lo que se espera, al poner en práctica el uso de metadatos, en el contexto de los cuadernos de laboratorio, es facilitar que los mismos estén disponibles para quien los necesite, sean transferidos, guardados y finalmente dispuestos acorde a los procedimientos aplicables a su clasificación. De igual

manera se espera que cualquier actividad experimental se presente de manera coherente para optimizar el intercambio del contenido informativo entre usuarios.

En el siguiente capítulo se hará énfasis en la importancia de añadir semántica a los datos o anotaciones hechas en diferentes tipos de procedimientos experimentales o técnicas de laboratorio, ya que es un aspecto estratégico cuando se está interesado en asumir las tecnologías Web emergentes, para mejorar el manejo del conocimiento. El uso de anotaciones normalizadas adoptadas de vocabularios controlados y ontologías, es una vía para facilitar el organizar, catalogar y describir la información disponible de tal forma que se pueda especificar el contenido y el contexto de la información, el propósito de la misma, indicar las relaciones entre los distintos datos, establecer quién es el autor/creador/propietario de la información y especificar la validez de la información.

3.7 BIBLIOGRAFÍA

BEATTY, P., DICK, S. & MILLER, J. (2008) Is HTML in a race to the bottom? A large-scale survey and analysis of conformance to W3C standards. *Ieee Internet Computing*, 12, 76-80.

CANTARA, L. (2006) Encoding controlled vocabularies for the Semantic Web using SKOS Core. *OCLC Systems & Services: International digital library perspectives*, 22, 111-114.

DASGUPTA, A. & RATH, D. S. (2009) Organizing Knowledge for Web Retrieval using SKOS: A Case Study in Human Protein Chain. *ICAL-Management Models and Framework*.

- DAVENPORT, T. H. & PRUSAK, L. (1998) *Working Knowledge: How Organizations Manage What They Know* Harvard Business School Press
- ERCEGOVAC, Z. (1999) Special topic issue: Integrating multiple overlapping metadata standards - Introduction. *Journal of the American Society for Information Science*, 50, 1165-1168.
- GRAY, A. J. G., GRAY, N. & OUNIS, I. (2008) Finding Data Resources in a Virtual Observatory Using SKOS Vocabularies *Sharing Data, Information and Knowledge*. Springer Berlin / Heidelberg.
- HARTMANN, J., PALMA, R., SURE, Y., SUAREZ-FIGUEROA, M. C., HAASE, P., GOMEZ-PEREZ, A. & STUDER, R. (2005) Ontology metadata vocabulary and applications. *On the Move to Meaningful Internet Systems 2005: Otm 2005 Workshops, Proceedings*, 3762, 906-915.
- HEERY, R. (1996) Review of metadata formats. *Program: electronic library and information systems*, 30, 345-373.
- KAKALI, C., LOURDI, I., STASINOPOULOU, T., BOUNTOURI, L., PAPATHEODOROU, C., DOERR, M. & GERGATSOULIS, M. (2007) Integrating Dublin Core metadata for cultural heritage collections using ontologies. *Dublin Core and Metadata Applications*.
- KANARE, H. M. (1985) *Writing the Laboratory Notebook*, Washington D.C, American Chemical Society.
- KIMURA, F., TOBA, T., TEZUKA, T. & MAEDA, A. (2009) Federated Searching System for Humanities Databases Using Automatic Metadata Mapping. *Dublin Core and Metadata Applications*

- MURPHY, L. D. (1998) Digital Document Metadata in Organizations: Roles, Analytical Approaches, and Future Research Directions. *31st Hawaii International Conference on System Sciences*
- NATIONAL AGRICULTURAL LIBRARY (NAL). & AGRICULTURAL RESEARCH SERVICE, U. S. D. O. A. (2010a) Glossary of Agricultural Terms. 2010 ed.
- NATIONAL AGRICULTURAL LIBRARY (NAL). & AGRICULTURAL RESEARCH SERVICE, U. S. D. O. A. (2010b) NAL Agricultural Thesaurus 2010 ed.
- OLSEN, S. (2004) IBM sets out to make sense of the Web.
- PAGANELLI, F., PETTENATI, M. C. & GIULI, D. (2006) A Metadata-Based Approach for Unstructured Document Management in Organizations. *Information Resources Management Journal*, 19, 22.
- PÄIVÄRINTA, T., TYRVÄINEN, P. & YLIMÄKI, T. (2002) Defining organizational document metadata: A case beyond standards. *Xth European Conference on Information Systems (ECIS)*. Gdańsk, Poland.
- STEINACKER, A., GHAVAM, A. & STEINMETZ, R. (2001) Metadata Standards for Web-Based Resources. IN PEIYA LIU, S. C. R. (Ed.) *IEEE MultiMedia*.
- UREN, V., CIMIANO, P., IRIA, J., HANDSCHUH, S., VARGAS-VERA, M., MOTTA, E. & CIRAVEGNA, F. (2005) Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Journal of Web Semantics*, 4(1), 15.
- WEIBEL, S. (2000) The Dublin core metadata initiative - The Frankfurt focus and the year 2000. *Zeitschrift Fur Bibliothekswesen Und Bibliographie*, 47, 3-13.